

THIS WEEK

EDITORIALS

HISTORY DNA discovery kicks off memorable blasts from the past **p.270**

WORLD VIEW Economic crisis triggering a modern Greek tragedy **p.271**



TICK TOCK The bacteria that keep time for the underwater squid **p.273**

Look after the pennies

Government decisions about where to spend and where to cut should be based on evidence, not ideology.

When a nation's expenses grossly exceed its income, as they routinely do in the United States, the most foolish way to curb the resulting deficit is to slash spending wildly. Yet that is the path that the US government has chosen to follow with this year's 'sequester'. A smarter way is to follow the path pioneered by evidence-based medicine: fund what works and cut what doesn't. That is the approach being pursued in the budget submitted last week by US President Barack Obama (see page 277).

In a budget chapter rarely mentioned in the media hoopla over proposed tax rises and spending cuts, the administration has laid out a blueprint to implement evidence-based decision-making throughout the government — in effect, bringing the methods of science to bear on policy.

This reform effort began under Obama's predecessor, George W. Bush, but it has accelerated greatly with the need to do much more with significantly less. The White House Office of Management and Budget (OMB) signalled the way that things were going in May last year when it instructed government agencies to incorporate evidence-based strategies throughout their operations.

The White House mandate applies to every agency in the executive branch, including those that fund science. But the most urgent target is the government's vast array of social services, which range from early-childhood enrichment projects to the home care of elderly patients. All were created with good intentions; few have had their effectiveness evaluated with any kind of rigour, or by anything that resembles peer review.

The OMB memorandum suggests several ways to change that situation. One is to fund social services through a 'tiered' approach not unlike the stages followed in clinical trials. At the lowest tier, agencies would allocate seed money to promising, but unproven, ideas, provided that the research programme builds in a rigorous assessment of outcomes by independent investigators — usually academic social scientists or non-profit research firms.

In the higher tiers, more funding would be available for programmes that are supported by stronger evidence (and with built-in assessment protocols). The highest tier would be reserved for large-scale, multi-million-dollar programmes that are already supported by multiple gold-standard, controlled trials.

Federal agencies are already using the tiered model for six evidence-based programmes — ranging from teen-pregnancy prevention to education — with budgets totalling about US\$1 billion in 2012. Obama's new budget proposes to boost that funding by 44% for 2014.

Another suggested strategy, pioneered in the United Kingdom, is a model known as 'social impact bonds' or 'pay for success'. It sees philanthropic organizations and private companies fund preventative services, with the government paying them back only if rigorous assessments show that the services save taxpayers' money. US federal agencies have tried this approach on a small scale with job-training programmes and projects to reduce recidivism in newly released

prisoners. In 2014, Obama proposes to spend up to \$195 million to expand these initiatives into areas such as housing and education.

Such strategies for data-driven decision-making have the potential to radically improve the US government's efficiency and effectiveness, and deserve vigorous support from Congress — with one caveat: both Congress and president must be equally vigorous about supporting research into what success actually means and how to measure it. These are still open questions in most areas of policy. Most parents, for

instance, probably think that there is more to a good education than getting their child to score well on standardized tests. It can be difficult to quantify such intangible benefits, but that is no excuse not to try.

Everyone favours government effectiveness as a concept. But every existing programme is also someone's livelihood. When those

judged ineffective — by whatever measure — are cut or consolidated, the protests and the lobbying are fierce. If officials can resist that pressure, evidence-based policy initiatives could help to bring about a much-needed shift in the inflamed fiscal debate, from ideology to pragmatism.

The OMB memorandum captured that spirit in a refreshingly un-bureaucratic call to arms: "Where evidence is strong, we should act on it. Where evidence is suggestive, we should consider it. Where evidence is weak, we should build the knowledge to support better decisions in the future."

That is easy to say; it is harder to do. But to say it is a start. ■

Smoke and mirrors

Italy's parliament must listen to expert advice before deregulating stem-cell therapies.

Just weeks after the white smoke from the Vatican signalled the election of a new pope, a grimmer pall hangs over the Eternal City — a fog of confusion and misrepresentation about stem-cell therapy. Those who have lit the fire beneath the debate say that they are promoting the translation of stem-cell research into the clinic so that currently incurable diseases can be treated. Nothing could be further from the truth.

The Second International Vatican Adult Stem Cell meeting, held on 11–13 April in Vatican City, was a shamelessly choreographed performance. Sick children were paraded for television, sharing the stage with stem-cell companies and scientists desperate to hawk a message that their therapies must be speeded to clinical use.

A kilometre away at the Italian senate, meanwhile, parliamentarians further eroded protection for vulnerable patients targeted by stem-cell companies. On 10 April, they amended an already controversial ministerial decree (see *Nature* 495, 418–419; 2013) with a clause that would redefine stem-cell therapy as tissue transplantation, thereby releasing it from any regulatory oversight. If the second parliamentary chamber endorses this amendment, Italy will be out of step with the rules of the European Union and the US Food and Drug Administration, both of which define stem cells modified outside the body as medicines.

Many scientists around the world were appalled by the events in Rome, and rightly so. It is wrong to exploit the desperation of the disabled and the terminally ill and to raise false hopes of quick fixes, as some at the Vatican meeting tried to do. It is also wrong to try to use such patients as experimental animals by bypassing regulatory agencies, as the Italian parliament seems to want to do.

Reputable stem-cell companies insist that stringent regulatory control is necessary, and that patients should be exposed to experimental treatments only when safety and efficacy is assured. Failures in the clinic will hold back the field. But not all of the cell-therapy industry is so tolerant.

With their ability to differentiate into various cell types, stem cells hold enormous potential to repair damaged tissues. Human embryonic stem cells can turn into any cell type, but many groups, including the Catholic Church, find their derivation from embryos unethical.

The current controversy concerns adult stem cells. These exist in several tissues, but can replace only those particular tissues. Big claims are being made for them, with many trials of therapies under way worldwide for conditions as diverse as Alzheimer's and heart disease. Some stem-cell therapies are approved by regulatory agencies; others sneak under the radar by exploiting rules allowing compassionate therapy, for example, or by operating in countries such as China or Mexico — and perhaps now Italy — where regulation is less strict.

The scientifically naive Vatican finds the concept of adult stem cells

attractive simply because embryos are not involved — yet it ignores the ethical implications of false hope.

The main organizer of last week's conference was the non-profit Stem for Life Foundation, launched by the stem-cell company NeoStem, both based in New York. The foundation says that it is in favour of strict regulation of stem-cell therapies. But its conference programme, which left no room for questions, included many speakers who clearly were not.

“A lot more research into the deep biology of stem cells is needed.”

It was framed as a fight for reason and fairness against an uncaring and intransigent scientific community.

Adult stem cells have already had clinical success, such as in bone-marrow transplantation for leukaemia treatment, growing new skin layers to treat burns and regenerating corneas. More ambitious hopes need to be tempered, however. Many trials involve infusing patients with mesenchymal stem cells from bone marrow, which are relatively easy to extract and grow. These can make only bone, cartilage and fat cells, but trials targeting other tissues use the rationale that other, non-stem properties of mesenchymal stem cells apply. It remains to be seen how effective these properties will be outside the normal biological home of the stem cells.

Given the burden of incurable disease, rapid bench-to-bedside translation is unquestionably crucial. But a lot more research into the deep biology of stem cells is needed. Some trials approved by regulatory agencies may yield useful results, but that is a long shot without strong research data. At least they are safer under regulatory eyes. Unregulated treatment — such as that issued on a compassionate basis by the Stamina Foundation in Brescia, Italy, which led to the current ministerial decree — is more worrying. The second parliamentary chamber needs to heed independent expert advice before voting to deregulate stem-cell therapies.

Stem cells will help to develop treatments for currently incurable diseases. But we are not there yet, whatever the smoke signals may say. ■

Due credit

Nature's podcast charts 12 landmark discoveries in the history of science.

Even the greatest scientific discoveries come with an element of the mundane. A humble paperclip was biophysicist Raymond Gosling's choice. Late one night in May 1952, in a chemistry lab in London, the PhD student wrapped DNA around a paperclip to keep the molecule's fibres stretched taut in front of an X-ray source so that he could analyse their structure. The result was the celebrated ‘photograph 51’ — the image that told James Watson that DNA strands curl around each other like a twisted ladder, and that the specific pairings in the rungs are key to the mechanism of inheritance.

The rest of that story is legend. Based on their work at the University of Cambridge, UK, Watson and Francis Crick published their paper in the pages of this journal, including a beautiful diagram of the double helix that was hand-drawn by Odile Crick, Francis' wife (J. D. Watson and F. H. C. Crick *Nature* 171, 737–738; 1953).

Next week marks the 60th anniversary of the publication of the famous Watson and Crick paper — and that of two other papers on DNA that appeared in the same issue. Neither was so high profile, but each was essential to the structure's discovery. Both were written by scientists at King's College London: one by Maurice Wilkins and his colleagues Alec Stokes and Herbert Wilson, and the other by Gosling and his PhD supervisor, Rosalind Franklin.

Only Gosling (now 86) and Watson survive from that group of seven scientists. Watson has never been shy, and his compelling swagger helped to establish another colossus of biology, the Human Genome

Project. But the supporting cast matters too, even on the biggest stages.

Gosling is a *Nature* author, even if he is largely forgotten when the story of DNA is told. To mark the anniversary of his paperclip-inspired contribution, *Nature* has interviewed him. You can hear the results at go.nature.com/lizfik, in the first of a series of monthly podcasts to highlight 12 key scientific discoveries from the pages of this journal. (Future episodes in the ‘Pastcast’ series will plunder the *Nature* archive to investigate the discovery of X-rays in 1896, the early days of quantum theory in the 1920s and the first report of the ozone hole in 1985.)

In the interview, a humble Gosling fondly recalls that Franklin's response to Crick and Watson's model of the double helix was gracious and sanguine: “She didn't use the word ‘scooped’. What she actually said was, ‘We all stand on each other's shoulders’.”

All three papers appeared with no peer review — unthinkable now. The head of the King's College biophysics unit, John Randall, belonged to the same London gentleman's club — the Athenaeum — as Lionel (‘Jack’) Brimble, co-editor of *Nature*. Randall convinced Brimble to publish Wilkins' paper alongside Watson and Crick's; Franklin's paper was added only after she petitioned for its inclusion.

This cavalier approach to submissions extended to the awarding of credit. Watson and Crick's paper features only a glancing concession to being “stimulated by a knowledge of the general nature of [Wilkins and Franklin's] unpublished experimental results and ideas”. There is no mention of Gosling by name. Gosling left research soon after, with no bitterness; in his words, he “was no good at it”.

Discoveries take ego, genius, conflict, inspiration and fierce ambition. But they also need the hard graft of PhD students who beaver away late into the night and improvise with what they find in the stationery cupboard. They do not always receive the recognition that they deserve. Raymond Gosling is a good place to start to reverse that trend. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Austerity-led brain drain is killing Greek science

Lack of funding and recruitment freezes are driving young researchers out of the country, warns Varvara Trachana.

Science in Greece is going backwards. This month, researchers lost access to the journal *Bioinformatics*, a top-ranked title in mathematical and computational biology. Many more publications are likely to disappear from Greek libraries. The Ministry of Education has not paid the bills for its subscription bundles. The largest publishers — including Elsevier, Springer and Taylor & Francis — have threatened to suspend access. Others have done so already.

The denial of scholarly papers, the lifeblood of research, to Greek scientists could mark the beginning of the end for creative science at universities and research institutes. We will no longer be able to keep up with international contributions. In areas such as biomedicine, it is crucial to have access to the latest information. Many Greek researchers, unable to afford personal subscriptions to their favourite journals, are already considering reviving a practice that was common a decade or so ago — contacting friends and colleagues in foreign research centres and asking them to fax or e-mail articles.

For many readers of *Nature*, the hardship faced by Greek scientists will come as no surprise. The country is reeling from six straight years of recession and unprecedented austerity measures. More than one-quarter of Greek people are out of work.

I am one of them. I am a biologist with a PhD in biological chemistry from the Aristotle University of Thessaloniki. In 2003, I went to Spain to work as a postdoc at the National Centre of Biotechnology in Madrid. In 2008, I returned to Greece as a research scientist with the National Hellenic Research Foundation in Athens, on a succession of short-term contracts. In March 2011, I was elected assistant professor of cell biology at the faculty of medicine of the University of Thessaly in Larissa. But I never began work there: I am one of about 800 faculty members who are still waiting to take up appointments around the country because the government refuses to approve the budget necessary for their salaries. They are distinguished scientists, many with years of postdoctoral experience, who have been selected through a long and demanding process and have been appointed by the heads of their respective universities.

The departments that selected these 800 faculty members are struggling to teach their students. In 2011, for the first time in decades, the Ministry of Education placed no new university professors. The scientific and professional prospects of young scholars in Greece are evaporating; this will leave the country's universities lifeless and impotent. Budgets for research institutes have been reduced by 30%. The 2013 education budget will cut funds by a further 14% and condemn Greece to scientific and educational dormancy.

There are no signs that the Greek government

understands that long-term commitment to funding science and education must be part of the strategy to boost economic growth. In 2007, even before the most recent cuts, university and research funding in Greece stood at 0.6% of gross domestic product in 2007, already far below the European Union average of 1.9%.

Wages of researchers and faculty members have been reduced by 20%. If someone tells you that the Greek economy fell because of giant public-sector salaries, tell them that the average monthly wage of a university lecturer here is now around €1,000 (US\$1,300). Researchers and professors who have spent years building their careers are asking whether it was worth it.

Scientists have adapted. They have learned to use the bare minimum of expensive reagents. They have become skilled at working with their overcoats on when deliveries of heating oil fail to materialize.

But many young scientists are heading abroad. In 2010, about 120,000 Greek scholars lived and worked elsewhere, about one-tenth of the total. The number is now estimated at 150,000. The young, skilled workforce, a key factor for economic development, is disappearing exactly when society needs it most. I, too, am considering whether to leave. The situation in Greece, combined with plans by European Union leaders to cut the research and development budget, paints a bleak picture for future generations.

Greek science is worth investing in and worth saving. In 2012, against all the odds, the proportion of the country's research that con-

tributed to the top 1% of most-cited articles was 13th in the world, above Canada, Italy and France. This human potential should be nourished by urgent government action.

A good place to start is to resist calls from existing professors to raise their retirement age from 67 to 70. The change has been suggested as a short-term fix for academic departments stretched to breaking point by the hiring freeze and retirements. Such a move would eliminate any chance that the new professors will be placed, given that the national programme of fiscal stability calls for the public sector to be dramatically reduced by making only one appointment for every ten retirements. It would block the natural and necessary renewal of university personnel with new blood. It would further accentuate the rapid ageing of Greek universities and leave faculties as shells containing only a handful of older professors and next to no lecturing staff and young researchers. We know from evolutionary biology that such micro-societies soon become extinct. ■

Varvara Trachana is an assistant professor of cell biology at the University of Thessaly in Larissa (appointment pending for two years). e-mail: vtrachana@gmail.com

HUMAN POTENTIAL
SHOULD BE
NOURISHED
BY URGENT
GOVERNMENT
ACTION.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/psqtsw

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROSCIENCE

Brain signature for thermal pain

Brain activity could one day help physicians to monitor pain, for which there is no reliable physiological test.

Using a type of magnetic resonance imaging that shows when certain parts of the brain are activated, Tor Wager at the University of Colorado Boulder and his colleagues started out by scanning the brains of 20 volunteers as they experienced warm to painfully hot sensations on their arm. They used these data to find a pattern of neural activation and deactivation that consistently appeared when the volunteers were exposed to painful heat. Further tests showed that this signature could discriminate physical pain from other stimuli, such as social pain and recollections of pain. The signature was reduced by analgesics. Such patterns could one day lead to more-objective assessments of pain. *N. Engl. J. Med.* 368, 1388–1397 (2013)

SYMBIOSIS

Roots spur on helpful biofilms

Components of plant cell walls can induce a bacterium that is used as a plant fertilizer to assemble itself into sticky mats known as biofilms.

Biofilms are often associated with hard-to-treat infections in animals, but those formed by the soil bacterium *Bacillus subtilis* actually protect plants from pathogens. A team led by Roberto Kolter at Harvard Medical School in Boston, Massachusetts, reports how

signals from the roots of the model plant *Arabidopsis thaliana* prompt bacteria to join up. Pectin and other polysaccharides on the surface of plant cell walls activate bacterial genes known to induce biofilms. The polysaccharides also form the raw materials that bacteria use to synthesize the extracellular matrix that holds the biofilm together. The plant, therefore, provides both the environmental cue and the building blocks

to promote beneficial colonization of its roots. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1218984110> (2013)

CONSERVATION BIOLOGY

Old evidence for fewer fish

The price of fish in the 1800s has helped to reveal the long-term effects of bottom trawling, a fishing practice in which nets are dragged across the seabed.

Ruth Thurstan, now at the University of Queensland in Brisbane, Australia, and her colleagues examined the testimonies of hundreds of fishermen in the northeast of England about changes to fish stocks and practices during the nineteenth century.

Statements about catch rate, price and fish size, which were given during two Royal Commissions of Enquiry (in 1863–66 and in 1883–85), revealed a perception by fishermen that numbers of



CLIMATE CHANGE

Desert plants reap no rewards

Atmospheres that are enriched with carbon dioxide can boost plant productivity in some ecosystems, but drought may prevent faster or greater growth in desert plants, such as those of the Mojave Desert (pictured) in the southwestern United States.

Researchers led by Beth Newingham then at the University of Nevada in Las Vegas measured productivity above and below ground for plants that were exposed to high levels of atmospheric carbon dioxide at experimental sites in the Mojave Desert for a decade.

They found that dominant shrub and grass species that were exposed to higher levels of carbon dioxide showed some gains in weight and photosynthesis in wet years compared with plants at control sites that were exposed to ambient carbon dioxide, but that these gains were not sustained during drought.

Desert ecosystems, which cover around one-third of Earth's land surface, may be more limited by water than by carbon, the authors suggest.

Glob. Change Biol. <http://dx.doi.org/10.1111/gcb.12177> (2013)

white fish, such as cod, had fallen by 64% during their careers. Many fishermen blamed the declines on trawling. These largely forgotten records suggest that trawling began to affect fish stocks long before scientific monitoring and official statistics began.

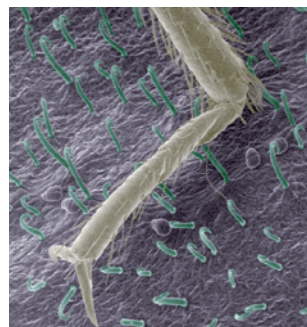
Fish Fish. <http://dx.doi.org/10.1111/faf.12034> (2013)
For a longer story on this research, see go.nature.com/5rbvcv

BIOMATERIALS

Leafy trap for bedbugs' legs

A natural Eastern European folk remedy catches bedbugs more effectively than synthetic mimics.

Catherine Loudon at the University of California, Irvine, and her colleagues used scanning electron microscopy and videography to document why bean leaves, which are traditionally scattered around an infected bed to trap bedbugs (*Cimex lectularius*) are effective. Tiny hairs on the leaves, known as trichomes, trap the insects by impaling their feet and hooking their legs (pictured). The researchers designed a synthetic mimic of the hairs with a polyvinylsiloxane moulding material often used in dentistry, and a hybrid mimic made of synthetic stalks and natural trichome tips. Although both mimics snagged the bugs, they could not impale them. This suggests that a better bug trap may need to recapitulate the mechanical properties of bean leaf trichomes more closely. *J. R. Soc. Interface* 10, 20130174 (2013)



MEGAN SZYNGLER/CATHERINE LOUDON

CELL BIOLOGY

Super-enhancers rule genes

The myriad biomolecules that regulate gene expression are governed by an unanticipated layer of control.

Researchers led by Richard Young at the Massachusetts Institute of Technology in Cambridge have found assemblies of enhancers — segments of DNA that associate with regulatory proteins and attach to genes to switch them on — that they have named super-enhancers. Compared with ordinary enhancers, super-enhancers bind to more proteins that activate gene transcription and are more sensitive to perturbation. The teams found that some cancer cells assemble pathological super-enhancers. Human cells contain tens of thousands of genes and many more enhancers, but most are controlled by only a few hundred super-enhancers that bestow characteristic properties on each cell type, the authors suggest.

Cell 153, 307–319; 320–334 (2013)

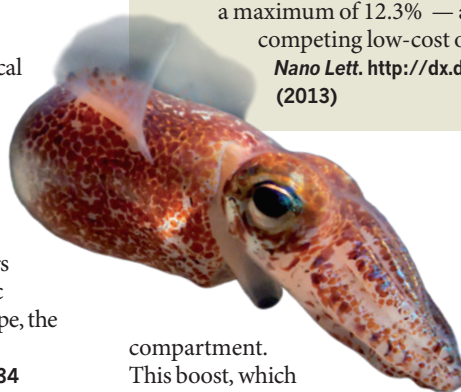
For a longer story on this research, see go.nature.com/spue5g

ECOLOGY

Symbionts set squid's clock

The tick-tock of an animal's daily clock can be set, in part, by beneficial bacteria.

At night, the bacterium *Vibrio fischeri* glows in a compartment in the underbelly of its symbiotic host, the Hawaiian bobtail squid (*Euprymna scolopes*; pictured), hiding the squid's shadow in the moonlight. Margaret McFall-Ngai at the University of Wisconsin—Madison and her colleagues have found that the light-producing bacteria also cause the squid to boost expression of its light-responsive gene *escr1* in the colonized



compartment. This boost, which helps to set daily rhythms, did not occur in bacteria-free squid, even if they were exposed to blue light to mimic bacterial glow. But when researchers provided squid with light as well as bits of the bacterium's cell envelope, *escr1* cycling returned. The researchers suggest that microbiota could also set daily rhythms in mammals. *mBio* 4, e00167-13 (2013)

IMMUNOLOGY

Proteins that rouse also quash

Signalling pathways that initially orchestrate potent antiviral attacks can, paradoxically, also help to sustain chronic infections.

Proteins called type I interferons (IFN-I) curb viral replication during early stages of infection, but

are unable to clear chronic infections. To solve this long-standing puzzle, independent groups led by David Brooks at the University of California, Los Angeles, and Michael Oldstone at the Scripps Research Institute in La Jolla, California, studied mice infected with strains of lymphocytic choriomeningitis virus that cause chronic infection.

Both groups found that IFN-I produces an initial burst of antiviral activity, but then depresses the immune system over the long term, allowing the virus to persist.

Chronic activation of the immune system can damage tissue, and the authors speculate that IFN-I signalling dampens the immune-system response to limit this damage. Inhibiting IFN-I signalling may help to control chronic viral infections, they say. *Science* 340, 202–207; 207–211 (2013)

► **NATURE.COM**

For the latest research published by Nature visit: www.nature.com/latestresearch

COMMUNITY CHOICE

The most viewed papers in science

PHOTOVOLTAICS

Cheap, colourful solar cells

★ **HIGHLY READ**
on pubs.acs.org
21 Mar–18 April

A light-harvesting crystal can make efficient, colourful and low-cost devices that convert sunlight into electricity.

Researchers led by Sang Il Seok at the Korea Research Institute of Chemical Technology in Daejeon, South Korea, created metal–organic hybrid materials that were composed of lead; methylammonium; and iodine or bromine, or both. Semiconductors used in solar cells can be coated in these materials, which absorb light and displace electrons that the semiconductor transfers to an external circuit. By varying ratios of bromine and iodine, the researchers tuned the material's stability and light-absorption properties to produce solar cells with a variety of translucent colours, from dark brown to yellow. On average, the cells could convert 10% of sunlight energy into electricity, with a maximum of 12.3% — a performance that matches competing low-cost options.

Nano Lett. <http://dx.doi.org/10.1021/nl400349b> (2013)

are unable to clear chronic infections. To solve this long-standing puzzle, independent groups led by David Brooks at the University of California, Los Angeles, and Michael Oldstone at the Scripps Research Institute in La Jolla, California, studied mice infected with strains of lymphocytic choriomeningitis virus that cause chronic infection. Both groups found that IFN-I produces an initial burst of antiviral activity, but then depresses the immune system over the long term, allowing the virus to persist. Chronic activation of the immune system can damage tissue, and the authors speculate that IFN-I signalling dampens the immune-system response to limit this damage. Inhibiting IFN-I signalling may help to control chronic viral infections, they say. *Science* 340, 202–207; 207–211 (2013)

SEVEN DAYS

The news in brief

POLICY

US budget proposal

President Barack Obama proposed nearly US\$143 billion for research and development in his 10 April budget plan, an increase of about 1% above 2012 levels. Applied research is slated to get \$1.8 billion more than basic science in the battle for funds. See go.nature.com/8i6bon and page 277 for more.

Global health plan

The United Nations Children's Fund and the World Health Organization on 12 April announced a US\$6.7-billion plan to treat and prevent childhood pneumonia and diarrhoea, which together kill 2 million children under the age of 5 each year. The Global Action Plan for the Prevention and Control of Pneumonia and Diarrhoea aims to cut deaths from these conditions by 67% and 95%, respectively, by 2025. It outlines 15 initiatives including wider vaccination, improved sanitation, provision of antibiotics and promotion of hand-washing.

Polio campaign

Leading scientists have supported a new global strategy to rid the world of polio by 2018. More than 400 researchers and public-health experts signed a

NUMBER CRUNCH

\$2.27 m

The amount paid at auction for Francis Crick's Nobel medal. The buyer was Jack Wang, chief executive of Chinese firm Biomob, which sells electromagnetic aids for healing. See go.nature.com/w43awo for more.



KATE HOLT/EYEVINE

Drought kills agricultural profit

Land degradation and desertification are causing huge reductions in agricultural profits in the world's poorest nations, according to a report presented at the United Nations Convention to Combat Desertification (UNCCD). The first assessment of its kind in more than 20 years calculates that Guatemala, for example, lost 24% of its agricultural gross domestic product to environmental degradation. The findings were presented at the UNCCD's second scientific

conference, on 9 April in Bonn, Germany. Nearly 3.7 million people in East Africa still require food assistance after droughts in Kenya (pictured) and other nations wiped out crops in 2011, the UNCCD says. "Desertification, land degradation and drought are key constraints to building social and environmental resilience, achieving global food security and delivering meaningful poverty reduction," said UNCCD executive secretary Luc Gnacadja at the meeting.

letter published on 11 April, endorsing plans to halt the spread of the virus in Nigeria, Pakistan and Afghanistan (the three countries never to interrupt transmission), and to integrate polio immunization with the delivery of other childhood vaccines. See page 290 for more on polio eradication.

RESEARCH

The costs of storms

Economic losses from powerful thunderstorms in the United States (those that cause more than US\$250 million in damages) doubled between 1970 and 2009, even when

normalized for socioeconomic changes and the growth of urban areas. Risk researchers with the insurance company Munich Re, based in Germany, presented the result at the meeting of the European Geosciences Union in Vienna last week. The increase mirrors the growing frequency of meteorological conditions that favour formation of thunderstorm cells, they say. See go.nature.com/qfltrq for more.

Telescope go-ahead

Hawaiian officials have given permission for the construction of the Thirty Meter Telescope (TMT) on

the 4,200-metre-high Mauna Kea mountain in Hawaii, project leaders announced on 13 April. Building could begin as early as April 2014. Thirteen telescopes already dot the sacred mountain, but the TMT would be the largest; the biggest optical instruments currently on Mauna Kea are the twin 10-metre Keck telescopes. See go.nature.com/cy6r5s for more.

BUSINESS

Publishing deal

The British start-up company Mendeley — a social academic network based in London, through which more than

MATT DUNHAM/AP

2 million users share and comment on research papers and references — has been bought by Elsevier, the Amsterdam-based science-publishing giant announced on 9 April. Neither company declared how much Mendeley had sold for, but observers familiar with the deal said that Elsevier had paid £45 million (US\$69 million).

Lab kit billions

A company that manufactures scientific instruments has announced one of the largest business deals of 2013 so far. Thermo Fisher Scientific, headquartered in Waltham, Massachusetts, said on 15 April that it will buy biotech-equipment manufacturer Life Technologies of Carlsbad, California, for US\$13.6 billion. The aim of the acquisition is to create a leading company in lab-kit production for research in proteomics, genomics and cell biology. See go.nature.com/27q9hf for more.

PEOPLE

IVF pioneer dies

Robert Edwards, Nobel-prizewinning pioneer of human *in vitro* fertilization (IVF), died on 10 April in Cambridge, UK, aged 87. Edwards (**pictured**) and his colleague Patrick Steptoe, who died in 1988, began research



into the technique in the late 1960s despite public criticism and tight research budgets. The first child conceived through IVF, Louise Brown, was born in 1978. Edwards was the sole recipient of the 2010 Nobel Prize in Physiology or Medicine for his work; Steptoe was not included because the prizes are not awarded posthumously.

Jewell confirmed

The US Senate voted on 10 April to confirm Sally Jewell, chief executive of the outdoor retailer Recreational Equipment, as the new head of the Department of the Interior, replacing Ken Salazar. As the department's chief, Jewell will oversee conservation, recreation, energy development and mining activities on more than 200 million hectares of public land and almost 700 million hectares offshore.

FUNDING

Space plans

Russia will give its Roscosmos space programme a massive US\$52-billion rise between now and 2020, President Vladimir Putin announced on 12 April. The first mission from its new Vostochny space port will be sent by 2015 with manned missions launching in 2018. 'Super-heavy' rockets capable of Moon missions are slated to be ready by 2020. See go.nature.com/grpejn for more.

EVENTS

Bird flu spread

The H7N9 avian flu virus has expanded its geographical range, with two human cases detected in Beijing, northern China, and two in the central Henan province. Previously, the virus had been limited to Shanghai and neighbouring regions on the eastern seaboard. As *Nature* went to press, 63 human cases including 14 deaths have been reported since China announced the first cases on 31 March. There is not yet any evidence of sustained human-to-human spread. See go.nature.com/bklq5s for more.

Romanian protest

All 19 members of Romania's National Research Council resigned on 12 April in protest at budget cuts and

COMING UP

23 APRIL

The US Institute of Medicine holds a public meeting in Washington DC to talk about priorities for research to reduce the threat of firearm-related violence.

go.nature.com/o5zump

22–25 APRIL

In Darmstadt, Germany, experts meet in a quadrennial gathering to discuss how to detect and clear up space debris — useless chunks of satellites and rockets that clutter Earth's orbit.

go.nature.com/uoyjtj

interference. The government had told the council how it should distribute basic research grants after slashing its budget on 8 April. Two days later, the education ministry launched a call for proposals in applied research, but under new rules removing a legal requirement for international peer review. The council says that such external review is necessary: in January, its respected peer-review system identified plagiarism in grant applications.

To steal a ship

A woman was detained by police in San Diego, California, on 9 April after apparently attempting to steal a research vessel belonging to the Scripps Institution of Oceanography. Members of the institute's science team were aboard the 52-metre-long *New Horizon* at the time but slept through the incident, says a Scripps spokesman. The vessel is now en route to Point Conception in California, where it will replace a climate and ocean-acidification monitoring station.

► NATURE.COM

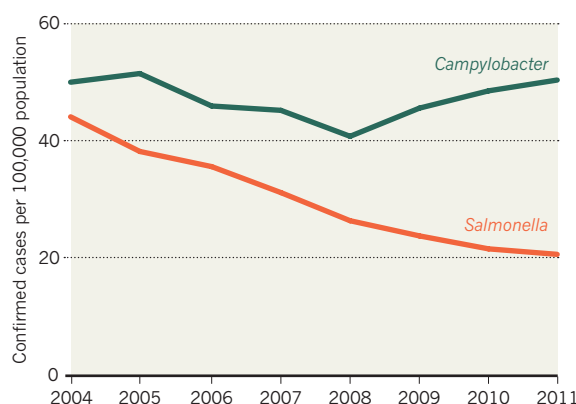
For daily news updates see:
www.nature.com/news

TREND WATCH

A vaccination programme to cut down on *Salmonella* bacteria in poultry is having success in Europe, where reported human cases fell for a seventh consecutive year, the European Food Safety Authority said on 9 April. But human disease caused by *Campylobacter*, the most common cause of food poisoning, is increasing: because infection does not cause disease in poultry, it is difficult to create a vaccine that generates an effective immune response against the bacteria.

BATTLE AGAINST FOOD-BORNE ILLNESS

Europe is cutting *Salmonella* infections — but making little progress in stopping *Campylobacter*.



SOURCE: EFSA/ECDC

NEWS IN FOCUS

SYNTHETIC BIOLOGY Can engineered organisms heal the planet? **p.281**

PHYSICS Quantum-mechanical weirdness survives a real-world test **p.282**

GENOMICS Coelacanth sequence reveals 'living fossil' genes **p.283**

PUBLIC HEALTH The hunt for polio's last stronghold **p.290**



ZHANG JUN/XINHUA/PHOTOSHOT



US President Barack Obama on 10 April unveiled a budget that proposes large increases for advanced manufacturing programmes.

POLICY

A back seat for basic science

Translational research wins in Obama's budget, but its economic value remains uncertain.

BY ERIC HAND, BETH MOLE, LAUREN MORELLO, JEFF TOLLEFSON, MEREDITH WADMAN AND ALEXANDRA WITZE

The magnolia trees behind the White House were decked in blossoms as US President Barack Obama announced his 2014 budget proposal on 10 April. They emphasized the lateness of the proposal, which is usually unveiled in early February. The short-lived flowers could also have symbolized a fragile optimism about the US economy. But with unemployment at a worrisome 7.6% and Congress insistent on deficit reduction, Obama's budget is cautious. The US\$3.8-trillion proposal would undo the

across-the-board cuts known as sequestration, but includes just \$143 billion for research and development, a mere 1% rise from 2012.

The budget reflects not only caution, but also the Obama administration's emerging attitude towards research: that science needs to be good for something, namely jobs and cures.

"To make America a magnet for good jobs, this budget invests in new manufacturing hubs," said Obama. "We'll spark new American innovation and industry with cutting-edge research like the initiative I announced to map the human brain and cure disease."

The first reference was to a one-time, \$1-billion request from the National Institute of Standards and Technology in Gaithersburg,

Maryland. The money would set up 15 manufacturing hubs, modelled after a pilot project devoted to three-dimensional printing that was established in 2012 in the former steel-manufacturing town of Youngstown, Ohio.

The second was to a multibillion-dollar brain-mapping project that would, in its first year, receive a total of \$110 million from the National Institutes of Health (NIH) in Bethesda, Maryland; the National Science Foundation (NSF) in Arlington, Virginia; and the Defense Advanced Research Projects Agency in Arlington. The project's first steps — creating tools to record the activity of millions of neurons — fall under basic research, but the White House has promoted the ▶

► project as key to understanding, treating and even curing brain diseases.

Those initiatives embody a spirit that runs throughout the proposal, reserving big boosts for research areas that promise near-term economic and social benefits (see ‘Tight times’). Adam Jaffe, an economist at Brandeis University in Waltham, Massachusetts, who specializes in science policy, says that the administration’s emphasis is natural: it makes funding science more politically palatable in a difficult economic climate. “They want to make clearer arguments as to why this is helping people in the relatively short run, rather than helping pointy-headed scientists,” he says.

But researchers unaccustomed to thinking about applications are starting to complain, says Jaffe. “They think it’s outrageous that the NIH should ask them to talk about the actual health-care implications of their research.”

Shifting money towards applications can leave basic research in the cold because the private sector will not support it, notes Barry Toiv, a spokesman for the Association of American Universities in Washington DC. Commercializing research discoveries can be a good idea, he says, but only if “the fundamental federal responsibility of funding basic research isn’t undermined”.

PHYSICAL SCIENCES

Under the budget proposal, NSF funding would rise by 8.4% over 2012 levels (used as a baseline because Congress did not finalize 2013 funding until late March). Much of that increase would go to purpose-driven initiatives such as the Cyber-Enabled Materials, Manufacturing and Smart Systems project, which would see its budget more than double, to \$300 million. The project funds research into new materials and robotics that could aid manufacturing.

The Innovation Corps (I-Corps), a training programme established in 2011 to guide researchers in turning their discoveries into products, was a pet project of former NSF director Subra Suresh, who stepped down last month. The programme’s budget would more than triple to \$25 million under Obama’s proposal. “I think this fulfils a need for the country,” says Dean Chang, associate vice-president of innovation and entrepreneurship at the University of Maryland in College Park, who was part of a team that won \$3.75 million from I-Corps this year to start a regional training centre. A dash of entrepreneurialism could allow “academia to contribute to the economic development of the country”, adds James Chung, executive director of the Office of Entrepreneurship at the George Washington University in Washington DC.

Engineering, with its applied focus, would get the largest boost of any research division at the NSF: 10.3%. That continues a trend: since the first Obama budget, for 2010, engineering funding has risen by 17%, to \$911 million. Yet

TIGHT TIMES

In US President Barack Obama’s budget request for 2014, applied research accounts for most of the increases in science funding (US\$ billions).

Agency	2012 actual	2013* estimated	2014 requested	Details
Biomedical research and public health				
National Institutes of Health	30.70	30.90	31.17	Nearly flat, but translational-science centre gets a 16% boost from 2012
Centers for Disease Control and Prevention	5.66	5.69	5.22	Includes \$10 million for research on the causes and prevention of gun violence
Food and Drug Administration	2.51	2.52	2.56	Overall \$4.7-billion budget includes more than \$2 billion in user fees
Physical sciences				
National Science Foundation	7.11	7.08	7.63	An 8% rise, with a robotic-manufacturing programme seeing a large gain
NASA (science)	5.07	5.12	5.02	Mission to capture an asteroid debuts, and funding to monitor near-Earth objects would double
Department of Energy Office of Science	4.93	4.90	5.15	Increases funds for applied and interdisciplinary energy research
National Institute of Standards and Technology	0.75	0.76	0.93	Agency is also pursuing a one-off request for \$1 billion to fund up to 15 manufacturing-innovation institutes
Earth and environment				
Environmental Protection Agency	8.45	8.50	8.15	A 3.5% decline from 2012 includes hits to research, as agency prioritizes cutting greenhouse-gas emissions
National Oceanic and Atmospheric Administration	4.91	5.26	5.45	A plan to maintain continuity in weather-satellite data would shift climate sensors to NASA
US Geological Survey	1.07	1.08	1.17	Boosts spending on ecosystem research and climate science

*2013 figures do not include the roughly 5% across-the-board cut arising from the sequester.

Source: White House Office of Management and Budget

the maths and physical-sciences division has barely gained at all, increasing from \$1.37 billion in 2010 to \$1.39 billion for 2014 (see ‘Apply thyself’).

BIOMEDICINE AND PUBLIC HEALTH

The NIH, ever popular with the public and with patients’ groups, often gets more from Congress than the president requests. For now, the administration is calling for a boost of just 1.5% over 2012 levels, to \$31.2 billion.

Projects that aim to speed treatments from labs to patients did much better. The fledgling National Center for Advancing Translational Sciences received a 16% boost to \$666 million — by far the biggest jump, proportionately, among the agency’s 27 institutes and centres. A big chunk of that — \$40 million — is an increase for the Cures Acceleration Network, which aims to make awards of up to \$15 million for work on ‘high need’ cures that industry is unlikely to develop.

The White House also returned to a problem that it first flagged up for special attention last year: Alzheimer’s disease, which cost the nation as much as \$215 billion in 2010, according to a study published on 4 April (M. D. Hurd *et al.* *N. Engl. J. Med.* 368, 1326–1334; 2013). The budget would add \$80 million to the roughly \$500 million that the NIH

is already spending on the disease.

“The NIH has always had this dual configuration as a science agency and as a health agency,” says Irwin Feller, a science economist at Pennsylvania State University in University Park, adding that the second function has expanded in recent years. The proportion of NIH money spent on basic research has fallen from two-thirds in the 1980s and 1990s to just over half now, says Matt Hourihan, director of research and development budget and policy at the American Association for the Advancement of Science in Washington DC.

Supporters of basic research “don’t particularly like” that trend, says Ben Corb, director of public affairs at the American Society for Biochemistry and Molecular Biology in Rockville, Maryland. He says that the NIH is feeling pressure from members of Congress, who “are asking: what are we getting for the \$30 billion we’re giving the NIH? Where are our cures? Where are our treatments?”

ENERGY

The Department of Energy (DOE) has always embodied the tension between basic and applied research. Its mission is to support applied work on energy extraction, even as its Office of Science funds blue-sky research that may not have applications for decades.

SOURCE: NSF

The White House request would boost the DOE's budget to \$28.4 billion, nearly 8% above 2012 levels, with the largest increase — almost \$1 billion — for initiatives targeting energy efficiency, renewables and advanced manufacturing. Clean energy “has been the signature of this administration”, says Michael Lubell, director of public affairs for the American Physical Society in Washington DC.

Investment would continue to flow to translational science at the Energy Innovation Hubs and Energy Frontier Research Centers initiated by departing energy secretary Steven Chu. The administration is also asking for a 38% rise in funding for high-risk work at the Advanced Research Projects Agency–Energy (ARPA-E), which was designed to bridge a ‘valley of death’ between basic and applied energy research. Support for the hubs and ARPA-E has faced resistance from Republicans in Congress, wary of picking winners in the applied sector. But Obama's administration hopes to launch prize schemes that, by providing incentives for innovation, may draw more bipartisan support. The budget includes \$200 million in a one-time ‘race to the top’ for states that provide the best proposals for investing in energy efficiency and advanced electricity-grid technologies, and a \$25-million prize for the first natural-gas combined-cycle plant to be equipped with carbon capture and sequestration technologies.

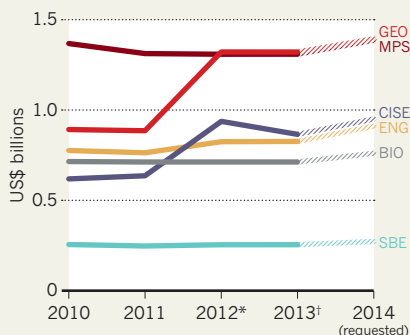
One of the largest increases in the Office of Science was reserved for basic energy sciences, a programme that, despite its name, supports applied ‘use-inspired’ materials research. Its budget would grow by 10% from 2012. Meanwhile, high-energy physics would suffer a 1.8% loss. With the shutdown of the Tevatron at Fermi National Accelerator Laboratory in Batavia, Illinois, in September 2011, the locus of activity in the field has shifted to the Large Hadron Collider at CERN, Europe's high-energy physics laboratory near Geneva, Switzerland. The budget cut will force officials at the DOE to make tough choices about remaining physics experiments, such as neutrino studies.

SPACE AND EARTH SCIENCES

NASA has been mission-oriented ever since it was tasked in 1961 with going to the Moon. Although some officials still advocate a return to the Moon (see *Nature* **492**, 161–162; 2012), this year, the president's budget proposal arrived on the same day as a congressional

APPLY THYSELF

Emphasizing applications, engineering has seen large gains at the US National Science Foundation.



*Computer science now includes funding formerly budgeted to the Office of Cyberinfrastructure, and Geosciences includes funding from the Office of Polar Programs. †Estimates do not include effects of sequestration. BIO: Biological Sciences; CISE: Computer and Information Science and Engineering; ENG: Engineering; GEO: Geosciences; MPS: Mathematical and Physical Sciences; SBE: Social, Behavioral and Economic Sciences.

hearing on the dangers of near-Earth asteroids, which seem to be the agency's current focus.

NASA wants to double the \$20 million spent annually on hunting dangerous space rocks, mainly to track smaller ones and to identify a candidate for a surprising mission. If Congress approves the administration's proposal, NASA would perform the first-ever asteroid-capture mission, in which a robotic spacecraft would snatch an asteroid 7–10 metres across, and tow it back to the vicinity of the Moon. Astronauts aboard the *Orion* crew vehicle could then visit and study the rock as early as 2021. The agency is asking for \$38 million to develop solar electric propulsion for the robotic part of the mission, and \$40 million to study how astronauts would work with the ‘uncooperative target’.

The Obama administration has emphasized NASA's Earth science, which would receive a 5% increase from 2012 levels. Other Earth-observing programmes, run by the National Oceanic and Atmospheric Administration (NOAA), also benefit in the request. Funding for NOAA's satellite division, which operates geostationary weather probes and the Jason-3 altimetry mission to monitor sea-level rise, would increase by 17%, to \$2.2 billion, outpacing the 11% increase for the agency overall.

Astrophysics and planetary science at NASA would suffer cuts of 1% and 19%, respectively, from 2012 levels. Yet those two divisions would still get big prizes: the James Webb Space Telescope, an infrared successor to the Hubble

Space Telescope that is on track for a 2018 launch, and a Mars rover to launch in 2020, built of spare parts from the Curiosity rover.

REALITY CHECK

As with all presidential budget requests, Obama's priorities are contingent on Congress. Mark Muro, a senior fellow at the Brookings Institution, a think tank in Washington DC, says that many of Obama's budget initiatives may amount to little more than “performance art” given the political focus on the deficit.

If the emphasis on commercially tinged basic research is to continue, science and technology economists agree that its value needs to be measured carefully. “It's a little hard to know how well it works, because we don't do any serious evaluation of these programmes,” says Jaffe.

That means measuring not just monetary returns on investments, but also the social return, says Julia Lane, an economist at the American Institutes for Research in Washington DC who until last year headed NSF efforts to measure research outputs. Negative results are part of the picture, she says. Private oil prospectors might not disclose locations that hold no oil, whereas the US Geological Survey would, which would benefit society. Similarly, unlike the NIH, drug companies may not report negative clinical-trial results.

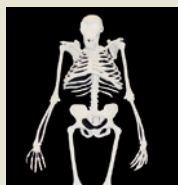
When mission-driven research is done at universities rather than in industry, it has an indirect social benefit: students who go on to found spin-off companies, says Lane. “The best way to translate knowledge is to wrap it up in a human being,” she says. “We need to capture that piece of information.”

Translational research can have pay-offs for basic science, says Matthew Begley, a mechanical engineer at the University of California, Santa Barbara, who, along with his colleagues, received \$1.2 million from the NSF to study high-temperature materials that could be used in aeroplane engines. He points out that the best scientists hop back and forth between basic and applied endeavours, and that basic research draws on applied science just as often as the other way around — witness the Large Hadron Collider, whose construction required major feats of engineering to allow it to discover the Higgs boson. “The line between science and technology is sort of blurry,” says Begley. “But when they're done well, the line is very blurry.” ■ **SEE EDITORIAL P.269**



MORE ONLINE

TOP STORY

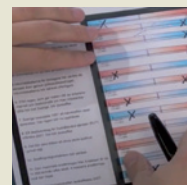


Australopithecus sediba proves to be hodgepodge of ape and human features
go.nature.com/wdasuq

MORE NEWS

- Egyptian mummies' DNA preserved despite hot climate go.nature.com/bkexht
- Supernova left iron-isotope signature in ancient bacteria go.nature.com/3hsqjc
- Lab-grown kidneys transplanted into rats go.nature.com/n3ub9p

VIDEO



‘Choice blindness’ trick transforms conservatives into liberals
go.nature.com/vr63qf

LEE R. BERGER/UNIV. WITWATERSRAND

LUND UNIV.

BIOENGINEERING

Synthetic biologists and conservationists open talks

But worries persist about unintended consequences of tinkering with nature.

BY EWEN CALLAWAY

Australian scientists made headlines last month when they revealed that they were close to cloning a frog, *Rheobatrachus silus*, last seen in the wild three decades ago. If they succeed, it may take another emerging technology to keep that frog alive.

Synthetic biology aims to endow organisms with new sets of genes and new abilities. Along with cloning, it has been portrayed in the press as a hubristic push to do fantastical things: bring back woolly mammoths or resurrect the passenger pigeons that darkened the skies of North America before they were eradicated by nineteenth-century settlers.

But at a first-of-its-kind meeting, held on 9–11 April at the University of Cambridge, UK, leading conservationists and synthetic biologists discussed how the technology could be applied in less fanciful ways to benefit the planet: to produce heat-tolerant coral reefs, pollution-sensing soil microbes and ruminant gut microbes that don't belch methane. Also on the list were ways to help frogs to overcome chytridiomycosis, the fungal disease threatening amphibians worldwide that is thought to have contributed to the extinction of *R. silus*.

The discussions took place against a background of mutual wariness, however. One synthetic biologist told the conference that he felt as though he were being treated like an irresponsible teenager who might accidentally wreck the planet. And conservationists bristled when a synthetic biologist announced that those in favour of the technology would win simply because they are younger than those who are against it. (He apologized the next day.) Kent Redford, a consultant for the Wildlife Conservation Society in New York and organizer of the meeting, made repeated calls for comity, plying attendees with free wine.

Such bickering is a sign that the fields are beginning to engage, says Drew Endy of Stanford University, California, who is widely seen as a co-founder of the synthetic-biology field.

"How are we going to explore the notion that we might aspire to rework our civilization such that it dances better with the planet? The synthetic-biology community, whatever the hell that is,



Sweet wormwood could be complemented by yeast as a source of the antimalarial artemisinin.

isn't going to figure it out by itself," he says.

Producing fuels, foods and medicines with microbes is the cutting edge of the field, stresses Richard Kitney, a synthetic biologist at Imperial College London. "We're not trying to produce woolly mammoths," he says.

SOIL SAVER

As an example of what the field can offer conservation, Kitney cites an undergraduate project he supervised that was presented at the 2011 International Genetically Engineered Machine competition, a kind of synthetic-biology science fair. Christopher Schoene, now at the University of Oxford, UK, and his team engineered *Escherichia coli* so that the bacterium would migrate into plant roots and produce the growth hormone auxin. In greenhouse tests, roots of cress plants that contained the engineered bacterium grew longer than those without, and the soil retained more water. Such a bacterium might help to combat desertification — the degradation of fertile land into desert when soil nutrients are lost.

But synthetic biology worries some observers, who fear what might happen if genes or organisms escape from their intended niches. Paul Falkowski, a geomicrobiologist at Rutgers University in New Brunswick, New Jersey, sees value in microbes that can turn carbon dioxide into fuel or make fertilizers from atmospheric nitrogen, but he worries that industrial-scale

production could have drastic consequences, such as the inadvertent production of greenhouse gases. "I am rather amazed at the naivety of synthetic biologists at the way the world works," he says.

Many attendees also expressed nervousness about the potential of synthetic biology to influence land-use patterns. Microbes that reduce greenhouse-gas levels might lessen the pressure on governments to maintain rainforests, they said. Technologies that make marginal lands more productive could turn undeveloped land into single-crop farms.

Such shifts are already beginning to occur. A project begun by Jay Keasling, a synthetic biologist at the Lawrence Berkeley National Laboratory in California, coaxed yeast to produce the antimalarial drug artemisinin at industrial levels (see *Nature* 494, 160–161; 2013). Much of the drug currently comes from cultivation of sweet wormwood (*Artemisia annua*), but Keasling believes that synthetic sources will eventually force *A. annua* growers in China and elsewhere to cultivate other crops. "I don't make the decision about what gets produced," says Keasling, whose company, Amyris in Emeryville, California, aims to produce industrial products with engineered microbes. "The marketplace decides. What I do is provide more options."

Concerns could be mitigated by designing ways to limit the spread of synthetic microbes. Schoene's team, for example, added a genetic safeguard to its *E. coli* that stops other microbes from acquiring the auxin-producing gene. "If [safeguards] are being developed with as much creativity as other technologies, that would reassure me a lot," says Stephen Palumbi, a marine biologist at Stanford University's Hopkins Marine Station in Pacific Grove, California.

Bill Sutherland, a conservation biologist at the University of Cambridge, agrees that his colleagues need to take synthetic biology seriously. But he says that a small poll he took at the meeting shows that the gulf between the two disciplines is not so wide. Both agree that more-efficient use of natural resources could be an important boon from synthetic biology. Both worry about the potential for synthetic organisms to harm natural ecosystems.

One issue that neither group was excited or very bothered about: restoring long-dead species. "It's a lot of fun," Sutherland says, "but it's not going to save the world." ■

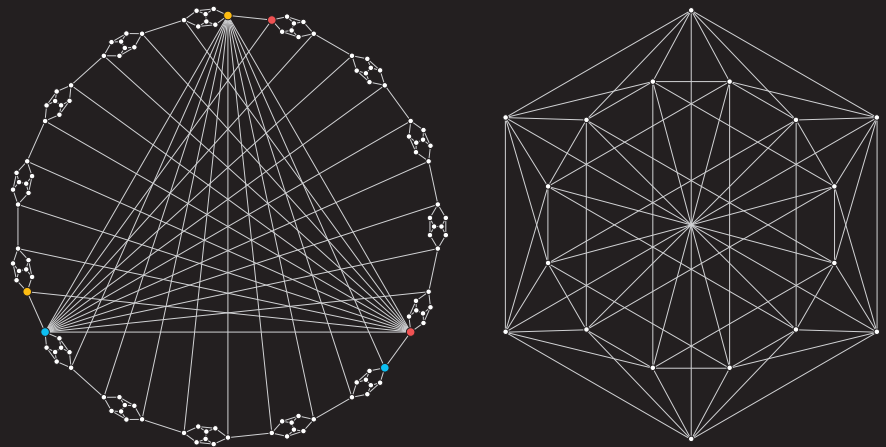
MICHAEL REYNOLDS/EPA

➔ **NATURE.COM**
For more on the
issues of synthetic
biology, see
go.nature.com/pls4ka

QUANTUM CONTEXTUALITY

Researchers have simplified the 117 contextual questions in the original 1967 Kochen–Specker theorem (left) to a more manageable 18 (right). Colours indicate duplicated questions.

- Yes/no question about quantum state of the system
- Connects questions whose answers cannot both be yes



QUANTUM MECHANICS

Photons test quantum paradox

Contextuality theorem could improve secure communication.

BY EUGENIE SAMUEL REICH

Like Schrödinger's cat, it was only supposed to be a thought experiment to elucidate the strange mathematics of quantum mechanics.

Now, the 46-year-old Kochen–Specker theorem, which describes the quantum dance of observer and observed, has passed its toughest test yet in the real world. The test, published in February (V. D'Ambrosio *et al. Phys. Rev. X* **3**, 011012; 2013), is indicative of growing interest in the theorem, triggered by new capabilities for manipulating photons and cold atoms (see 'A quantum revival').

"We can test things that until now were just mathematics," says Adán Cabello, a physicist at the University of Seville in Spain, and a co-author of the paper. "We've been waiting for the technology." Although his team has focused on the pure maths of the theorem, follow-up work may eventually find practical use in defending encrypted conversations against attack, and in improving random-number generators.

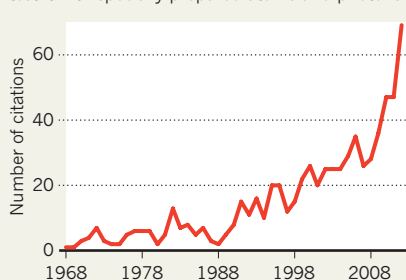
The theorem, first published in 1967 by mathematicians Simon Kochen and Ernst Specker, shows that it is incorrect to assume that the results of a quantum mechanics experiment are determined before measurements are made (S. Kochen and E. P. Specker *J. Math. Mech.* **17**,

59–87; 1967). That assumption is valid in classical physics; for example, the heat content of a cup of tea is unaffected by the thermometer measuring it. But it breaks down in quantum mechanics, where measurements change their subjects in ways that depend on what else is being measured — as if a set of thermometers conspired to create the heat that they measure.

This behaviour is called quantum contextuality. One example of the principle is 'spooky action at a distance', in which the quantum states of two particles are entangled such that measurements on one particle instantaneously influence the other, even if it is far away.

A QUANTUM REVIVAL

Citations of the 1967 Kochen–Specker theorem have soared since physicists have been able to test it with specially prepared atoms and photons.



Testing contextuality for particles that are not entangled has been hard because the Kochen–Specker theorem is so complex: it considers the answers to a set of 117 measurements on a single quantum particle that can exist in three or more quantum states. Preparing photons or atoms in more than two quantum states has been possible only in the past few years.

The latest demonstration is the most faithful validation of the theorem yet, because it tests a set of 18 measurements that are mathematically very similar to the 117 described by Kochen–Specker (see 'Quantum contextuality'). Cabello and his colleagues manipulated the angular momentum and polarization of identical photons to create different combinations of four quantum states, then watched how the photons produced interference patterns in a detector. The researchers confirmed that the outcomes of the measurements are coordinated, just as Kochen–Specker implies has to be the case for quantum mechanics to hold.

Specker died in 2011. Kochen, an emeritus mathematician at Princeton University in New Jersey, is impressed with the demonstration. "We thought it was a *Gedankenexperiment*, a thought experiment," he says.

Tests of the theorem might have practical uses. Contextuality holds promise for cryptography, says Cabello, because it implies that a spy eavesdropping on messages sent through quantum systems will influence measurements by the agents trying to communicate. If the agents find disruptions to the contextuality of their messages, they can halt the communication.

Similarly, the results of quantum contextual measurements can generate random numbers, which are also valuable in cryptography, says Luming Duan, a physicist at the University of Michigan in Ann Arbor who has done simplified Kochen–Specker tests on cold atoms. Only pure quantum measurements are truly random, and Kochen–Specker tests can verify that no classical effect has interfered with the random-number generation, he says. "Contextuality provides a way to do this without using entangled states", which are delicate and easily destroyed.

Some physicists have doubts about the technological applications. Adrian Kent, a quantum theorist at the University of Cambridge, UK, says that Kochen–Specker measurements are difficult to perform, and that an unknown classical aspect of a measurement could upset the quantum measurement. That loophole could be used to hack encryption based on the Kochen–Specker theorem, he notes.

Fabio Sciarrino, a physicist at the Sapienza University of Rome and a co-author of the latest paper, says that his group has quantified the extent of this loophole in its test. He is confident that the test is precise enough to prevent classical objects from masquerading as quantum ones. "We found very high agreement between experiment and theory," he says. That, at least, keeps the strangeness of quantum mechanics on a sure footing. ■

SOURCE: KOCHEN AND SPECKER (1967); D'AMBROSIO ET AL. (2013)

SOURCE: ISI WEB OF KNOWLEDGE

GENETICS

‘Living fossil’ genome unlocked

The genes of an ancient fish, the coelacanth, have much to reveal about our distant past.

BY CHRIS WOOLSTON

The South African fisherman who pulled a prehistoric-looking blue creature out of his net in 1938 had unwittingly snagged one of the zoological finds of the century: a 1.5-metre-long coelacanth, a type of fish that had been thought to have become extinct 70 million years earlier.

Since then, scientists have identified two species of coelacanth, one African and one Indonesian. With their fleshy, lobed fins — complete with bones and joints — and round, paddle-like tails, they look strikingly similar to the coelacanths that lived during the Cretaceous period, when dinosaurs still roamed Earth.

Now, an international team of scientists has sequenced and analysed the genome of the African coelacanth, *Latimeria chalumnae*; the findings are reported on page 311.

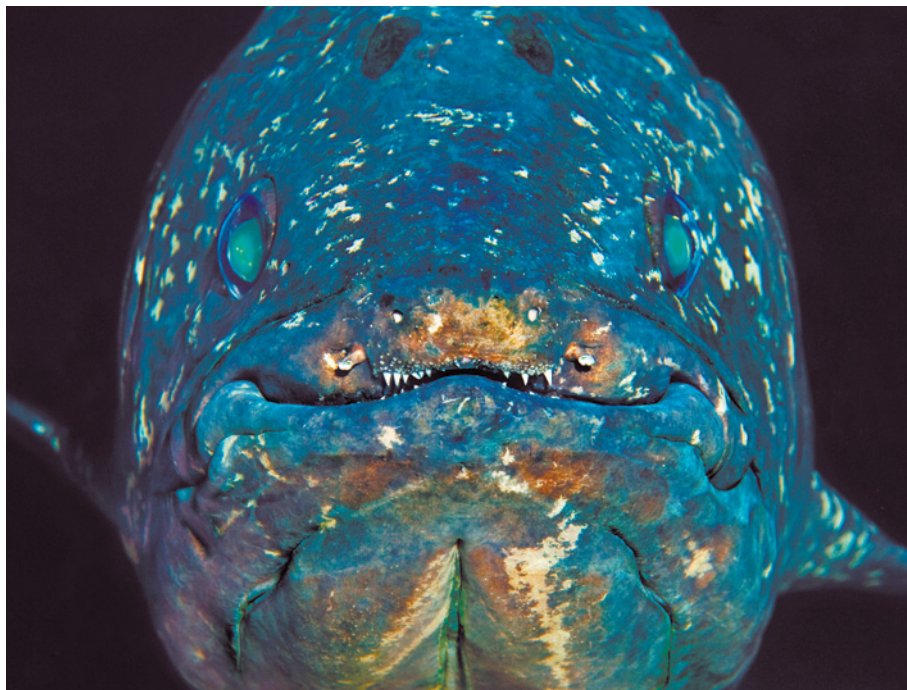
Like lungfish, the other surviving lineage of lobe-finned fishes, coelacanths are actually more closely related to humans and other mammals than to ray-finned fishes such as tuna and trout. Ancient lobe fins were the first vertebrates to brave the land, and the coelacanth genome is expected to reveal much about the origins of tetrapods, the evolutionary line that gave rise to amphibians, reptiles, birds and mammals, says lead author Chris Amemiya, a biologist at the University of Washington in Seattle. “The coelacanth is a cornerstone for our attempt to understand tetrapod evolution,” he says.

Ending one long-standing argument, analysis of the coelacanth genome clearly shows that it is not the closest living fishy relative to tetrapods, Amemiya says: that honour belongs to the lungfish. However, he adds, the lungfish genome is unlikely to be sequenced any time soon because it is much larger and more complicated than that of the coelacanth.

Although coelacanths are often called “living fossils”, these fish haven’t really been frozen in time, says co-author Kerstin Lindblad-Toh, a comparative genomicist at Uppsala University in Sweden. Comparison of protein-coding genes in coelacanths with those of cartilaginous fishes shows that the coelacanths have been steadily accruing DNA changes.

But the rate of change has been remarkably slow. The latest analysis shows that the genes of modern coelacanths can themselves be

“The coelacanth is a cornerstone for our attempt to understand tetrapod evolution.”



LAURENT BALLESTA

The African coelacanth is a close relative of the first fishes to venture onto land.

considered living fossils, says James Noonan, a geneticist at Yale University in New Haven, Connecticut.

Scientists already had hints of the coelacanth’s sluggish evolution. In a 2012 study, researchers in Japan and Tanzania compared the DNA of the African and Indonesian coelacanths. Specifically, they looked at *HOX* genes, which help to guide embryonic development (K. Higasa *et al. Gene* **505**, 324–332; 2012). Even though the two species separated, by one estimate, perhaps 6 million years ago, their genes are remarkably similar. For these particular genes, the difference between the two species of coelacanth was about 11 times smaller than that between the *HOX* genes of humans and chimps, two species that parted ways perhaps 6 million to 8 million years ago.

SLOW TO CHANGE

It is impossible to say for sure, but the slow rate of coelacanth evolution could be due to a lack of natural-selection pressure, Lindblad-Toh says. Modern coelacanths, like their ancestors, “live far down in the ocean, where life is pretty stable”, she says. “We can hypothesize that there has been very little reason to change.” And it is possible that the slow genetic change explains why the fish show such a striking resemblance to their fossilized ancestors.

The analysis showed that not all parts of the coelacanth genome are slow to evolve. The genome has a large number of transposable elements — non-coding parts of the genome that have an important role in gene regulation — that have been moving around in the genome at a relatively rapid pace. Non-coding DNA could be a significant source of evolutionary change, Lindblad-Toh says. But Amemiya adds that, for now, the role of non-coding DNA in speciation is “speculative”, and its significance in coelacanth evolution is not clear.

As expected, the genome holds clues to the genetic changes behind the transition from a lobed fin to a tetrapod limb, Amemiya says. For example, the analysis found that coelacanths and tetrapods share a regulatory gene sequence that helps to promote limb development. But other findings came as a complete surprise. The fish is the first vertebrate found to lack genes for immunoglobulin-M, an almost universal immune-system protein. Instead, it has two genes for a distantly related immune protein that evidently “pick up the slack”, he says.

Further analysis of the genome is bound to reveal much about our own distant past, adds Noonan. “It will allow us to identify the genetic drivers of tetrapod evolution, the genes and regulatory elements that are responsible for the vertebrate land transition.” ■

GLOBAL WARMING

Climate models fail to 'predict' US droughts

Simulations identify past megadroughts, but at wrong times.

BY QUIRIN SCHIERMEIER

“This would be a fine country if it only had water,” observes a settler looking at the barren west Texas plains. “So would Hell,” replies a despairing farmer.

That old Texas joke probably originated in the 1950s, when the state was baked by its most relentless drought in recorded history. Last year, rain kept clear of the region again, and scientists predict that the entire North American southwest will become increasingly drought-prone as climate change proceeds¹.

Reliable forecasts of future ‘megadroughts’ would be a boon to farmers and water managers. But results presented last week at the annual assembly of the European Geosciences Union in Vienna suggest that such forecasts are still beyond the reach of current climate models.

Sloan Coats of Columbia University’s Lamont-Doherty Earth Observatory in Palisades, New York, and his colleagues tested whether a state-of-the-art climate model could simulate the droughts known to have occurred in the southwest during the past millennium. The model incorporated realistic numbers for factors that affect temperature and rainfall, such as atmospheric carbon dioxide levels, changes in solar radiation and ash from volcanic eruptions.

It also incorporated changes in the El Niño/Southern Oscillation (ENSO), a recurring temperature anomaly in the tropical Pacific that greatly affects weather in the western United States and many other parts of the world. (The warm phase — El Niño — often brings torrential rain and flooding; the cold phase — La Niña — tends to bring drought.) The team then compared the results of its simulations with data from the North American Drought Atlas, a detailed compilation of droughts based on the thickness of tree rings.

The results were puzzling. Although the simulation produced a number of pronounced droughts lasting several decades each, these did not match the timing of known megadroughts. In fact, drought occurrences were no more in agreement when the model was fed realistic values for variables that influence rainfall than when it ran control simulations in which the values were unrealistically held constant. “The model seems to miss some of the dynamics that drive large droughts,” says study participant Jason Smerdon, a researcher at Lamont-Doherty who studies historical climate patterns.

Other climate models tested by the team fared no better, he says. In particular, the models failed to reproduce a series of multi-decadal droughts that occurred in the southwest during the Medieval Climate Anomaly, a period between AD 900 and 1200 when global temperatures were about as high as they are today.

The problem may lie in the models’ inability to reproduce the cycling between the ENSO’s El Niño and La Niña phases, especially given that many scientists think that La Niña is the major driver of drought in the southwest. The ENSO “behaves much messier in the real world than in climate models”, says Jessica Tierney, a climate scientist at the Woods Hole Oceanographic Institution in Massachusetts who has investigated the role of the ENSO in East African rainfall variability². “We’re not sure how it

“The model seems to miss some of the dynamics that drive large droughts.”

has varied in the past, and we don’t know how it might change in response to climate change. This is really one of the big uncertainties we’re facing.”

In addition to their failure to reproduce El Niño and La Niña, existing models do not fully capture other factors that influence rainfall, such as clouds and vegetation. But Smerdon adds that the atmospheric and oceanic dynamics that inhibit rainfall and favour prolonged drought may be essentially random and so almost unpredictable.

Last week’s findings highlight the broader challenge of predicting how precipitation patterns will change as the global climate warms. Models are often at odds over the very direction of regional changes. For example, different projections prepared for the Colorado Water Conservation Board disagree on whether mean precipitation in the state will increase or decrease by 2050 (ref. 3).

But the uncertainties don’t change the larger picture, scientists say. “Climate models are not perfect, but they do the big things really well,” says Tierney. “We can be pretty confident that the southwest will warm and that water will become scarcer.” ■

1. Seager, R. *et al.* *Science* **316**, 1181–1184 (2007).
2. Tierney, J. E., Smerdon, J. E., Anchukaitis, K. J. & Seager, R. *Nature* **493**, 389–392 (2013).
3. Ray, A. J. *et al.* *Climate Change in Colorado: A Synthesis to Support Water Resources Management and Adaptation* (Colorado Water Conservation Board, 2008); available at go.nature.com/xs8471.

SPLINTERS OF THE AMAZON

Decades after Thomas Lovejoy isolated fragments of the Brazilian rainforest in a grand experiment, researchers are building on his legacy around the world.

BY
JEFF TOLLEFSON

Ecologist Thomas Lovejoy tucks his trousers into his socks with a casual warning about chiggers and then hikes off into the Amazon jungle. Shaded by a tall canopy and dense with ferns and underbrush, the old-growth forest looks healthy, but Lovejoy knows better. Three decades ago, the surrounding forest was mowed down and torched as part of a research project, and the effects have spread like a cancer deep into the uncut area. Large trees have perished. The spider monkeys have moved out, as have the army-ant colonies, and many of the birds that depend on them.

Lovejoy and his team have been studying this 10-hectare fragment of forest since the late 1970s as part of the largest and longest-running experiment in tropical ecology. In collaboration with ranchers, they cleared the trees around this and ten other plots of varying size to create islands of intact forest. The researchers have been monitoring the plots ever since, documenting how deforestation harms the adjacent untouched forest as specialist plants and animals gradually give way to generalists and pioneer species that prefer disturbed habitat. "We are chronicling the simplification of these forests," says Lovejoy, a professor at George Mason University in Fairfax, Virginia.

Covering roughly 1,000 square kilometres in an area north of Manaus in the central Amazon, the experiment was set up to test fundamental theories about the viability of small, disconnected ecosystems. By documenting pervasive changes in the forest fragments, Lovejoy and his co-workers provided the first hard data that conservationists needed to promote the preservation of extensive areas of intact forest. "It's the most important ecological experiment ever done," says Stuart Pimm, a conservation ecologist at Duke University in Durham, North Carolina, who has collaborated on the project. "We knew that small and isolated was bad, but we needed to know how bad."

The researchers are now exploring the long-term effects of habitat fragmentation, but the ecological record there is ironically threatened by forest that is taking over abandoned pastures. Although Lovejoy has struggled to maintain financing for long-term monitoring, the US National Science Foundation is breathing new life into the project by funding the team to isolate some of the plots anew.

The experiment has also helped to train and inspire a

generation of 'fragmentologists', who are working around the world to understand the cascade of ecological impacts that follow human development. Most notably, in early April, an international team started chopping down trees in Borneo as part of an nearly £6-million (US\$9-million) experiment that replicates and extends the Brazilian one.

"The Amazon experiment changed the game," says Rob Ewers, principal investigator on the Borneo project at Imperial College London. "I like to think of our project as the next step."

THE AMBASSADOR

"Welcome to Camp 41," says Lovejoy, beaming at a group of guests he invited to tour the experiment — and do a little bird-watching — over New Year's Eve, an annual tradition. Fit at 71, he has a slight paunch, a crop of thinning hair and pale skin that is a touch reddish from the heat and the hike to his forest base. Lovejoy offers a quick tour of the open-air shelters that house hammocks and dining facilities as well as the bathrooms, showers and a makeshift pool down a trail by the stream. Over the years, he has entertained a long list of high-profile guests here, ranging from Al Gore (when he was a senator) to actor Tom Cruise and high-ranking Brazilian officials.

Lovejoy has always served as a scientific ambassador and chief fund-raiser, and left the fieldwork to others. After cleaning up from a day tramping around the forest, he sits beneath a cashew tree and begins plying his guests with caipirinhas, the national cocktail of Brazil. Peering over wire-framed glasses, he guides conversations about the strange beauty of tropical creatures, environmental policy and the history of science and development in the Amazon. Darkness falls, and an orchestra of frogs claims the forest.

Lovejoy arrived in the Amazon to study birds as a graduate student from Yale University in New Haven, Connecticut, in 1965, just as concerns about development in the region were rising among scientists and politicians.

That same year, Brazil enacted its modern Forest Code, which at the time required ranchers and farmers in the Amazon to maintain a 'legal reserve' on 50% of their land (the legal reserve is now 80%).

➤ NATURE.COM

To see a slideshow about the Brazil experiment, visit: go.nature.com/e48opz



MARK MOFFETT/MINDEN PICTURES/NATIONAL GEOGRAPHIC STOCK

Two years later, entomologist Edward O. Wilson and biologist Robert MacArthur published their influential *The Theory of Island Biogeography*, which laid the foundations for the modern understanding of species diversity and rates of extinction in isolated habitat, whether surrounded by water or by agricultural fields.

Soon after Lovejoy earned his PhD in 1971, ecologists became embroiled in what became known as the SLOSS debate, which stood for 'single large or several small'. The question was whether it would be better to protect massive continuous landscapes or many smaller biodiversity hotspots. Lovejoy thought about the Brazilian law and realized that the legal reserve could provide a way to probe these questions. "The ranchers were going to clear the land anyway," he says. "My crazy idea was that maybe you could arrange the 50% and create a giant experiment."

The project kicked off in 1978, with \$500 a month from conservation group the WWF and the support of the Brazilian National Amazonian Research Institute (INPA) in Manaus. Lovejoy hired another former student from Yale, ecologist Rob Bierregaard, to run the project. A year later, he and a team of Brazilian scientists began surveying the forests areas they were planning to isolate, which came in sizes of 1, 10 and 100 hectares (see 'Fractured forest'). The first wave of machetes and chainsaws came through in June 1980, and in September, Bierregaard's team walked the perimeter of the plots, dripping burning rubber onto forest debris. When the conflagration died down, the first two square patches of old-growth rainforest were left standing amid swathes of smoking embers that remained hot enough to cook the crew's beans for days.

The early phase of the experiment was hardly smooth. The Brazilians complained that it was too much of a US

initiative, and the ranchers were slow to clear the rest of their land. One year passed, then two. "It was really frustrating," says Bierregaard, who is now at the University of North Carolina at Charlotte. "We were publishing totally unreplicated results from the 1- and 10-hectare reserves."

By 1983, rather than waiting for the ranchers, Lovejoy secured more funding from the WWF to create another pair of fragments. The results began rolling in immediately, with the edges of the plots showing a substantial loss of key species. Yet, as the experiment grew, Lovejoy's team soon had more data than it could deal with, and in 1996 he brought on Bill Laurance to help make sense of the plant data. In 1997, Laurence and the team reported that up to 36% of the biomass in the first 100 metres of the forest fragments had disappeared in 10–17 years of isolation¹. "It really taught people how edge effects are driving rapid changes in ecology," says Laurance, who is now stationed at James Cook University in Cairns, Australia.

The main drivers are sunlight and air circulation. As the pastures and forest edges heat up each day, the air over those regions rises, drawing cool moist air out of the forest. The hot dry air takes a toll on large hardwood trees such as mahogany and ebony. The open fields also expose the forest to wind, which blows down trees and further opens up the canopy. Over time, these gaps are filled with fast-growing trees and vines. These pioneer species eventually seal off the forest like a scab, helping to delay further impacts, but neither the carbon density nor the diversity of the forest recovers quickly. Today, the researchers continue to track these effects as they work their way through the forest.

Those early results suggested that scientists were underestimating the broader impact of fragmentation, and in 1998, Laurance extrapolated the findings across the

'Islands' of forest near Manaus, Brazil, are allowing researchers to see how deforestation affects the local ecology.

"WE KNEW THAT SMALL AND ISOLATED WAS BAD, BUT WE NEEDED TO KNOW HOW BAD."

A FAMILY OF FRAGMENTS

A pioneering study in the Amazon spawned nearly two dozen experimental studies of isolated habitats in other parts of the world.



tropics. His team's calculations suggested that the biomass loss around forest edges could produce up to 150 million tonnes of carbon emissions annually² — exceeding emissions from the United Kingdom.

In 2003, Lovejoy and his fellow fragmentologists took their first stab at one of the questions that had originally inspired the project: how big is big enough? Documenting a 50% decline in the number of bird species living beneath the canopy in the 100-hectare plots over the first 15 years of isolation, they derived a kind of half-life for extinction: the time it takes to lose half of the species increases roughly tenfold with a 1,000-fold rise in the area of a reserve³.

The calculations were based only on birds and may not capture the ecological dynamics of larger reserves. But for Lovejoy, they suggest that a reserve would need to be on the order of 100,000 hectares in area to maintain relatively stable levels of biodiversity. Given that it is impractical in most places to set aside such large areas, this suggests that most fragmented landscapes around the globe may be doomed to a continuing decline in biodiversity.

Yet, the fragment experiment also points to a possible solution: the secondary forests that have sprung up in abandoned pastures. The regrown forests created wildlife corridors, which allowed army-ant colonies, birds and howler monkeys and other small mammals to migrate from intact forest into the experimental plots. For Pimm, those findings served as a call to action. In 2005, he created a non-profit conservation organization called SavingSpecies, with a goal of identifying biodiversity hotspots such as Brazil's Atlantic rainforest that would benefit from habitat corridors — a kind of lifeline to the remaining forest.

"This is where the science comes together with the strategic and tactical conservation efforts," Pimm says. In 2007, his group and its partners purchased a ranch that separated a small population of endangered golden lion tamarins from a larger forested area. After the team moved out the cattle and replanted some areas, secondary forest started to grow. Last year, the researchers saw the first evidence of golden lion tamarins and cougars moving through former ranchland. "Our bet on studying forest fragments paid off," he says.

Looking forward, Lovejoy says that many crucial questions remain about the processes playing out in the fragmented ecosystems. How many of the species in the plots are doomed to extinction? How quickly and deeply

will these impacts move through the fragmented forests? Will rapid shifts in insect and other animal populations drive long-term changes in seed dispersal and thus plant diversity? And what role will global warming have?

Continued monitoring of the fragments could produce some answers, but the secondary growth around the plots has undermined the viability of the experiment. The US National Science Foundation provided money for a new bout of clearing later this year, which is part of a larger \$450,000, five-year grant to continue bird research there. "This grant will extend our record to nearly 40 years," says Phil Stouffer, an ornithologist at Louisiana State University in Baton Rouge and principal investigator on the grant. For Lovejoy, it is a new lease on life for the experiment.

As the project continues, long-term data from the fragments and from the larger control plot in the adjacent primary forest could be invaluable for studying the impacts of global warming, says Adalberto Val, who heads the INPA, which provides part of the half-million-dollar or so annual operating costs of the experiment. The rest comes from the Smithsonian Tropical Research Institute in Panama and various foundations. He calls the forest-fragments project a "scientific treasure".

GLOBAL SPREAD

In 1984, a team led by Robert Holt, now at the University of Florida in Gainesville, implemented one of the first follow-on studies, focusing on experimentally designed patches of secondary regrowth amid the agricultural fields of Kansas (see 'A family of fragments'). By 1990, another four projects were under way, and a survey published in 2003 identified a total of 21 fragmentation experiments of various scales and durations⁴. Andrew Gonzalez, an ecologist at McGill University in Montreal, Canada, has replicated Lovejoy's experiment at the bench- and chamber-scale with isolated patches of moss; he has used the results to inform his own work on wildlife corridors in Montreal⁵.

The experiments have produced a range of results, but most of them confirmed significant impacts along the edges of the fragments that broadened over time. And the subsequent studies have reinforced the findings that habitat corridors can help to sustain the isolated fragments.

Lovejoy's latest scientific progeny — third-generation fragmentologist Ewers — initially worked in the Amazon as a 25-year-old postdoc under Laurance. Ewers first visited

Camp 41 in 2004 while conducting unrelated research on deforestation. He is now setting up a similar experiment in the state of Sabah in Borneo. Whereas Lovejoy's experiment focused on primary old-growth forest, Ewers wanted to observe the evolution of landscapes that have already been degraded and transformed by humans — something more representative of modern environmental realities. He set up fragments in logged forests and palm plantations as well as plots within forest reserves to investigate the ecological and conservation value of different landscapes.

"The general idea of Lovejoy's project was just breathtaking 30 years ago," says Ewers. "But we have new questions today, and I think we are in a much better place to make sense of these data in the long run."

His initiative, known as the Stability of Altered Forest Ecosystems (SAFE) project, is receiving some £6 million pounds in core funding over ten years from the Sime Darby Foundation in Kuala Lumpur, the philanthropic arm of one of the world's largest palm-oil producers. Forty-two plots mirroring the sizes used in the Amazon project will be located at various distances from the surrounding forests. Dozens of scientists have been surveying the sites, and the project now has data on 3,000 to 4,000 tropical species, from ants and beetles to birds, bees and trees — a database much larger than the one Lovejoy and Bierregaard were able to compile at the outset of their experiment. After a year's delay, loggers fired up the chainsaws on 4 April.

The unpublished baseline data show a landscape severely affected by development. For instance, the average amount of carbon locked up in trees in the primary forests has been estimated at 243 tonnes per hectare, compared with 49 tonnes per hectare in logged forest and just 4 tonnes per hectare within palm plantations. And yet, while many logged forests have been hard-hit ecologically, they still show pockets of remarkable biodiversity, including all five of the native cat species.

Ed Turner, an ecologist at the University of Cambridge, UK, who helped to set up the experiment, says that the project is designed to inform discussions about how to conserve habitat in a region dominated by palm-oil production. "The lessons from SAFE have the potential to have a very far-reaching effect," he says.

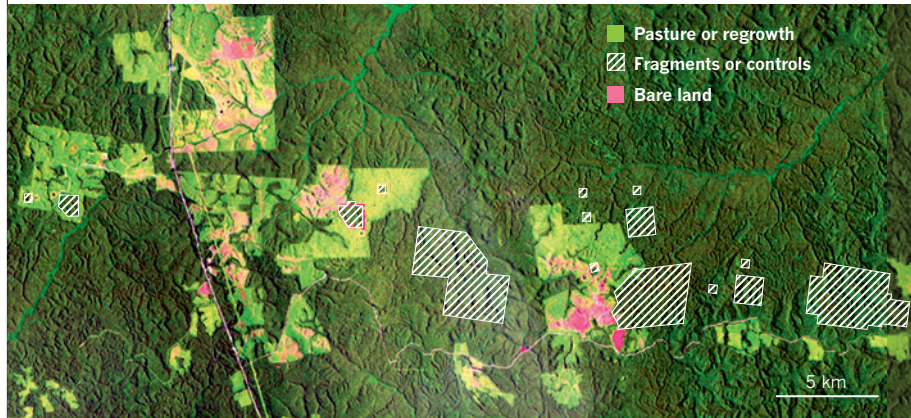
Another component of the Borneo project will focus on how nitrogen and other nutrients move through soils and plants, with the goal of understanding the biogeochemical processes at work throughout the forests as they are affected by fragmentation. The UK National Environmental Research Council is expected to announce in the coming months nearly £5.7 million for a series of projects focused on the relationship between biodiversity and biogeochemical cycles. For example, initial surveys have suggested that leaf litter decomposes roughly twice as fast in logged and primary forests as it does in palm-oil plantations. Once scientists have isolated fragments in each of these systems, they will be able to track changes in the decomposition rates as edge effects take hold, then relate those changes to soil nutrient cycles and impacts on plant communities.

"It's about doing fundamental science to help us understand what happens to a forest when you change it in these ways," says Ken Norris, an environmental biologist at the University of Reading, UK, and biodiversity theme leader for the UK National Environmental Research Council. "And if you are going to do ecosystems science, you need experiments outdoors at the scale at which these systems work."

Surveying the ongoing work, Lovejoy is pleased to see

that his crazy idea still has legs. The project in Borneo will advance the field, perhaps leading to new collaborations and comparative studies with researchers at his own experiment in the Amazon. "I started out thinking this would be a 20-year experiment, and then I would get the answer and we would be done," he says, gazing into the rain one afternoon at Camp 41. "It turned out to be more complex. So many new questions arose."

In the end, he says that much of the project's impact — and his own — has come from broader educational and advocacy efforts that have advanced conservation initia-



FRACTURED FOREST

The Biological Dynamics of Forest Fragments Project in Brazil has experimental plots of primary forest ranging in size from 1 hectare to 100 hectares located within cleared pastures.

tives in Brazil and around the world. He notes that Brazil has now protected nearly half of its share of the Amazon, more than scientists like himself could have hoped for several decades ago. He also points to the hundreds of Brazilian scientists who have moved through the project, including people such as Rita Mesquita, a researcher at the INPA who first came to the project as a student in 1985 and rose up to become scientific coordinator and an important state environmental official.

Lovejoy says that the scientific value of the Amazon fragment experiment only increases with time. Basic questions about the rate of species extinctions in fragmented habitat still plague the field of ecology, and the fragments provide a unique way to explore the issue. Ultimately, he envisages raising enough money to buy the ranchland surrounding the project and then convert it into an educational facility for science and ecotourism, and he has a new advisory board that is looking into options.

"If we can get this place stabilized and institutionalized, we could do a lot of things," he says. One way or another, he seems confident that the research will continue. "Even if my plane goes down on the way to Miami, I think it will happen."

Lovejoy falls silent, and the murmur of rainfall envelopes Camp 41. After a pause, he heads for the hammocks to join his guests for an afternoon nap. Soon enough, the rains will stop and Lovejoy will once again don his safari hat and binoculars, and set off to explore the fragmented forest. ■

Jeff Tollefson covers energy and environment for *Nature* from New York.

1. Laurance, W. F. *et al.* *Science* **278**, 1117–1118 (1997).
2. Laurance, W. F., Laurance, S. G. & Delamonica, P. *For. Ecol. Manage.* **110**, 173–180 (1998).
3. Ferraz, G. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 14069–14073 (2003).
4. Holt, R. D. & Debinski, D. M. in *How Landscapes Change* (eds Bradshaw, G. A. & Marquet, P. A. pp 201–223 (Springer, 2003).
5. Staddon, P., Lindo, Z., Crittenden, P. D., Gilbert, F. & Gonzalez, A. *Ecol. Lett.* **13**, 543–552 (2010).



Polio's moving target

Finding and vaccinating Nigerian nomads may be one of the last obstacles to the eradication of polio.

BY EWEN CALLAWAY

Mohammed Abubakar's home is not on any map — at least not yet. To reach his settlement in a desolate part of northern Nigeria, four health workers creep over deep-rutted roads in an old Peugeot for an hour, then ride motorcycles over narrow dirt trails for another 30 minutes — stopping only for the odd herd of cattle. Finally, they spot a cluster of mud-brick huts, known to the Fulani nomads who live there as a ruga.

“*As-salamu alaykum*,” — peace be with you — says Ardo Babangida, a traditional leader accompanying the team. Children swarm around the visitors, and Daniel Santong, an easy-going veterinarian and leader of the group, asks to meet Abubakar, the head of the household. Meanwhile, a young colleague whips out

a smart phone and uploads the settlement's Global Positioning System coordinates into a database. Abubakar arrives, clad in a lavender tunic and white skull cap, and Santong tells him that they are trying to eliminate polio in nomadic people. Abubakar clasps his guest's hands in appreciation. He says that he cannot remember the last time that health workers came to vaccinate his children. It is a story that Santong and his colleagues are now accustomed to hearing, even though door-to-door immunization campaigns happen on a near-monthly basis in the region.

These dusty paths are the front lines of polio eradication. A 25-year,

US\$10-billion global effort has taken the number of polio cases from hundreds of thousands per year to just hundreds, but it is now struggling to stamp the virus out of its final strongholds in Pakistan, Afghanistan and Nigeria, where transmission has never been interrupted. Of these, Nigeria was the only one to see an increase in cases from 2011 to 2012, and public-health experts worry that the virus's recalcitrance here will prevent global eradication, and eventually lead to a wider resurgence of the disease.

The barriers to polio eradication in Nigeria are complex and numerous. The country does not have a working public health-care system, and some local government officials are less

Vaccinators reach a child in Bauchi, Nigeria.

RUTH MCDOWALL

► NATURE.COM
To watch a video report on this story, visit:
go.nature.com/evovt3

than committed to the cause. In urban centres in the north, widespread distrust of the government leads many parents to refuse vaccination for their children. What is more, in February several polio workers were murdered — for unknown reasons — at health clinics in Kano, northern Nigeria's largest city.

But epidemiologists have identified one barrier that might be overcome cheaply and safely: locating and counting remote populations, including the nomadic livestock herders who drift through the region with the changing seasons. Records of their numbers and movements are incomplete, but the population is thought to include hundreds of thousands of young children, many of whom have received none or only some of the multiple oral polio vaccine doses required to achieve full protection. Proponents of the programme say that nomads are a polio reservoir, spreading disease around the country during their migrations. So in June 2012, the National Stop Transmission of Polio (N-STOP) programme, organized through the Global Polio Eradication Initiative (GPEI) and supported by the Nigerian government, started a census of Fulani nomads and other hard-to-reach populations, as part of a global emergency action plan against polio.

"Until we solve the problem of these unvaccinated nomads, we're not going to fix polio," says Frank Mahoney, a veteran field epidemiologist leading the project from Abuja. "We're not going to be able to eradicate it."

ROADBLOCKS AHEAD

On a scorching, cloudless December day, a procession of several hundred livestock has taken over a road running between the states of Bauchi and Kaduna in northern Nigeria. Men walk alongside their cows and sheep, while women and children creep along on motorcycles. It is the height of the dry season, the landscape is parched, and the group is heading south to graze its livestock. Here, a local government area (LGA) in Kaduna is travelled by a sizeable population of Fulani pastoralists like these during their biannual migrations. It is also one of dozens of 'high-risk LGAs' on which N-STOP teams are focusing their efforts. No polio cases have yet been detected here, but here is not far from the borders of Kano and Bauchi, which both recorded cases in the past year.

Mobile and remote populations are often strongholds for disease. Somali nomads contained some of the final cases of smallpox, and the vaccination of herds in remote patches of east Africa was crucial to eradicating the cattle disease rinderpest, completed in 2011.

In Nigeria, Fulani nomads receive little education or health care from the government. "Nobody looks after them, nobody takes primary health care to them, nobody remembers they exist," says Endie Waziri, a member of one of the N-STOP teams. When polio vaccination workers do visit their remote settlements, she says, they tend to visit only the first ruga they



Many Fulani migrate twice a year, making it difficult for vaccination programmes to find them.

see and not look for others in the vicinity.

Nigeria made significant headway against polio after starting an eradication programme in 1996. But those gains were erased in 2003, when Muslim clerics in the northern state of Kano called for a boycott of the polio vaccine over fears that the eradication campaign was a Western conspiracy to sterilize the population. Soon, Kano, Kaduna and other northern states had halted all polio vaccination campaigns. The boycott ended a year later, but by then polio had exploded across northern Nigeria and started to seep into nearby countries, such as Cameroon and Côte d'Ivoire, that had previously vanquished the virus. Nigeria has since made lurching progress against the disease. Cases fell from more than 1,000 in 2006 to 21

"If we hit these areas, we get a much bigger bang for the buck," says Chima Oluabunwo, a Nigerian epidemiologist who took a sabbatical from his position at Morehouse School of Medicine in Atlanta, Georgia, to serve as the field coordinator of the nomads project.

Still, the lack of a working public health-care system has been a problem, says Heymann, who points out that neighbouring countries with large nomadic populations such as Chad have successfully interrupted transmission. "Other countries with migrant populations have done the job," he says. India, for instance, offered polio vaccinations at train stations to catch migrants, and Chad offers combined veterinary and polio vaccination services to encourage Fulani nomads — who depend on

"Other countries with migrant populations have done the job."

in 2010, before rising again to 122 last year. Many more cases probably went undetected.

Religious opposition to vaccines among settled populations has now given way to refusals driven by disenchantment. "People want things other than polio vaccination," says David Heymann, chairman of the advisory board for Public Health England and the former head of the polio-eradication efforts for the World Health Organization (WHO). "They can't understand why people are coming once a month to give them vaccination when what they want are treatments for their children with fever or diarrhoea."

The nomads, however, rarely refuse polio vaccination for their children, and they are eager to receive other health and veterinary services.

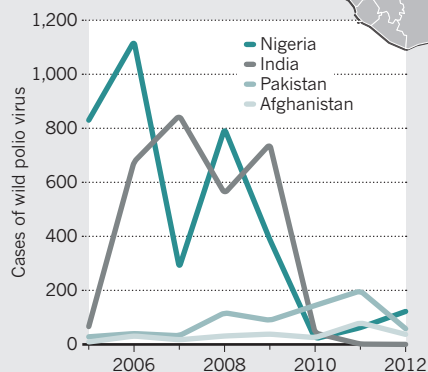
the health of their livestock — to take part.

N-STOP's census programme was designed to support on-going vaccination efforts, uncovering areas of need and directing resources and local vaccination teams to them. In fact, so as not to interfere with local efforts, N-STOP teams didn't bring vaccine stocks with them until the federal government asked them to. Since August, the N-STOP surveys have uncovered more than 32,000 settlements and identified more than 700,000 children — nearly 40,000 of whom had never been vaccinated against polio.

Although little more than 3% of the 122 polio cases reported last year in Nigeria occurred in nomad children, the teams discovered more than 100 probable cases that went unreported,

Last holdouts

In 2012, five countries worldwide reported cases of wild polio virus (as opposed to cases related to a vaccine strain). Sustained transmission continues only in Pakistan, Afghanistan, Nigeria and until recently, India. Only Nigeria saw the number of cases rise between 2011 and 2012.



Mistrust of the government in Nigeria's northern states leads to high rates of vaccine refusal, and nomadic children are often missed by vaccinators.

Viruses from Nigeria spark short-lived outbreaks across other parts of Africa. Chad recorded 132 cases in 2011, but just 5 in 2012.

India, now officially polio free, last recorded a case in January 2011. Neighbouring Pakistan and Afghanistan still serve as reservoirs for outbreaks.

Number of reported cases:
 ■ 0 ■ 1-50 ■ 51-100 ■ 101-150

SOURCE: GLOBAL POLIO ERADICATION INITIATIVE

supporting the idea that nomads form an important link in the chain of transmission. When not on long migrations, Fulani nomads interact with other people at markets. More than one-third of confirmed polio cases in 2012 were among children who lived in close proximity to nomadic communities. And nomad migration routes are hotspots of low vaccination coverage, according to data from the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. The surveys are “really a strategy that’s been missing in the tool box for a long time, and it needs to be urgently done”, says Mahoney. But Mahoney also notes that the project is a work in progress. Emmanuel Musa, the WHO coordinator in Lere, is not sure that it will be feasible for his district’s vaccination team to reach all the nomad communities that N-STOP teams are uncovering.

“If we hit these areas, we get a much bigger bang for the buck.”

“There are inadequate funds,” he says.

At a catch-up in December, Fururatu Zakari, the WHO coordinator for Kaduna state, also questioned whether the mapping data will be enough to allow health workers to locate nomadic communities. “They should be showing teams where settlements are and not just writing it down,” she tells Olaniran Alabi, the programme’s field coordinator for Kaduna. Zakari complains that the N-STOP programme isn’t coordinating its day-to-day activities with local health authorities, which could result in confusion and duplicated efforts.

The sheer scale of the GPEI initiative may

explain some of the tensions. The public-private effort, which includes the WHO, CDC, the United Nations Children Fund (UNICEF), Rotary International and the Bill & Melinda Gates Foundation, has become the world’s costliest public-health initiative, and one of the longest running. Since it started in 1988, it has missed three deadlines for halting transmission — in 2000, 2005 and 2012 — and it now burns roughly \$1 billion per year chasing the last remaining pockets of disease.

CAUSE FOR OPTIMISM

Organizers are optimistic about winning the war, however. India was long thought of as the Waterloo of the initiative because its high population density and poor sanitation provided ideal conditions for the virus to spread. But it celebrated its second year without a

the state’s governor, Rabiun Musa Kwankwaso.

Counting and vaccinating nomads will not solve all Nigeria’s polio troubles, but it is easier to achieve than tackling domestic terrorism, vaccine refusal and other challenges.

When asked how N-STOP will measure the success of the nomad programme, Mahoney puts it simply: “Stopping polio transmission. That’s the big indicator.” But he also points to the thousands of settlements and children that have been mapped and counted so far.

As *Nature* went to press, Nigeria had recorded 11 cases of polio this year; it had 17 in the same period last year. President Goodluck Jonathan has vowed to bring that number to zero before his term ends in 2015. Michael Galway, senior programme officer at the Bill & Melinda Gates Foundation in Seattle, Washington, is optimistic that the push to reach remote populations is paying off. “The work of the nomads project has been extremely beneficial in opening the eyes of the [polio eradication] programme to this additional piece of the puzzle,” he says.

The Nigerian project could hold lessons for future efforts to eradicate disease (measles is on some agendas), which are likely to hinge on reaching mobile and remote populations.

Targeting vaccination efforts on nomadic children is the right strategy, says Paul Rutter, a spokesman for the independent monitoring board set up to evaluate global efforts at polio eradication in 2010. But it will be the people behind it who ultimately dictate its success. “This kind of dogged determination to reach every last child will be what rids Nigeria of polio.” ■

Ewen Callaway writes for *Nature* from London.

COMMENT

EVOLUTION We must not forget the lessons of Willi Hennig, father of cladistics **p.295**

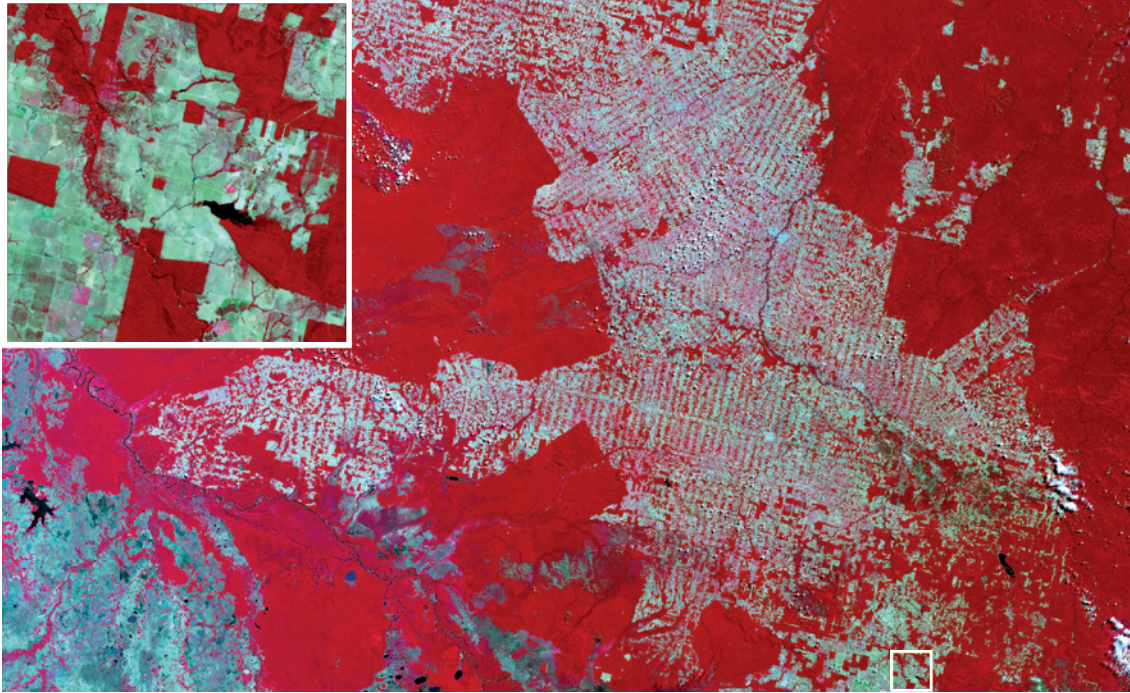
CAREERS E. O. Wilson's pragmatic, passionate advice to young scientists **p.297**

EXHIBITION Show of rare treatises charts history of anaesthesia **p.298**



PUBLISHING What sorts of science books are selling well? **p.299**

UK-DMC2/DMC INTERNATIONAL IMAGING



An image from the UK-DMC2 satellite shows forest in the Brazilian Amazon in red and cleared areas in green (detail in inset).

Choose satellites to monitor deforestation

Illegal logging threatens tropical forests and carbon stocks. Governments must work together to build an early warning system, say **Jim Lynch** and colleagues.

Tropical deforestation contributes 12% of total anthropogenic carbon dioxide emissions globally¹. Illegal logging is costing nations tens of billions of dollars each year. Although governments are making headway on agreements to stop this destruction, so far there is no coherent plan to monitor tropical forests on the scale or timescales necessary to do so.

Incentives are being negotiated for states to implement the United Nations REDD+ framework: Reducing Emissions from Deforestation and Forest Degradation, extended to include conservation, sustainable management of forests and the

enhancement of forest carbon stocks. The Intergovernmental Panel on Climate Change (IPCC) is also developing forest remote-sensing plans for consideration by the 19th Conference of the Parties (COP-19) to the United Nations Framework Convention on Climate Change (UNFCCC), which will be held in Warsaw this November.

Satellites provide the only means of viewing vast forest areas regularly — the tropics cover almost half of Earth's land area. But basic decisions have yet to be made on which Earth-observing systems should be used and how forest data should be monitored, reported and verified. In our view, the mapping strategies

proposed so far are too sparse and slow, making it impossible to identify forest damage until at least a year later.

We believe that an early warning system is needed to allow authorities to stop illegal logging quickly. Two strategies are necessary to achieve this: first, a new set of tropical orbiting radar satellites that can 'see' through clouds to monitor global forests daily; second, a plan for existing satellites to assess forest carbon stocks several times a year, to account for seasonal variations.

The REDD+ working group, which meets in Bonn, Germany, from 29 April to 3 May, must agree on a comprehensive, ►

► rapid-response global monitoring system in time for it to be ratified at COP-19.

REAL-TIME MONITORING

More than one billion people in the tropics depend on forests for their livelihood. The illegal logging trade is estimated to be worth between US\$30 billion and \$100 billion annually, with governments losing \$10 billion each year in tax income². Stolen wood is estimated to depress world timber prices by up to 16%.

As well as curbing these losses, implementing REDD+ and halving deforestation rates by 2050 would prevent the release of 50 petagrams of carbon in total into the atmosphere^{3,4}, saving a potential warming of 0.14 °C. This is a small but crucial step towards potentially much larger carbon savings.

Implementing REDD+ will require many facets of the carbon cycle to be monitored, including tree cover and carbon emissions from felling. There is also a need to assess the state and health of forests by analysing colour changes due to chlorophyll loss. Biomass, dead organic matter and soils must be included in the carbon budget.

Because different types of forest vary in their capacity to store carbon, changes in land use must be recognized. For example, Indonesia's policy of converting its 91 million hectares of tropical forest to oil palm plantations reduces the land's value from a REDD+ perspective, because much less carbon is being stored. Around 72% of the frontier forest has already been lost.

Optical and radar satellites working in different parts of the electromagnetic spectrum can monitor all of these characteristics. But international bodies and governments have been slow to formulate and agree on common guidelines for so doing. Without the right technologies or globally accepted assessment standards, billions of dollars could be wasted on projects that do not deliver.

The European Union has focused on preventing illegal logging, rather than specifying how forestry information should be collected by countries signed up to its 2005 Forest Law Enforcement, Governance and Trade regulation. This has led to a proliferation of incomparable collection methods. Moreover, there is no plan for REDD+ carbon credits to be accepted in the European Union Emissions Trading Scheme.

Some groups have published guidelines for satellite monitoring for REDD+, including the UN Food and Agriculture Organization's Global Observation of Forest and Land Cover Dynamics programme and the Group on Earth Observations' Global Forest Observations Initiative. In our view, these reports lack ambition and an understanding of the potential of satellites. Another group, the Committee on Earth Observation Satellites, for example, foresees doing only one optical



Illegal logging in Indonesia costs billions.

image survey a year — insufficient to track seasonal variations in forest carbon stocks.

To achieve good annual coverage, optical measurements should be taken at least every 1–2 weeks. To spot illegal logging, coverage needs to be even more frequent. Observations must be made at least daily, with data analysed much more quickly than at the current rate of months. If we are to protect tropical forests, we must go beyond mapping damage long after it has happened and instead offer early warnings.

WAY FORWARD

Governments and the IPCC need to accept that satellites are the only efficient and realistic way to provide monitoring for REDD+^{5,6}. This must be enshrined in international law through the UNFCCC process and documented in detail. Governments must commit to securing and maintaining systems of Earth-observing satellites to perform the monitoring.

Remote-sensing satellites operate in two spectral regimes: optical and radar. Optical sensors in different colours are sensitive to vegetation greenness, fractional tree cover, forest type and vegetation density. They can cover millions of square kilometres in a single image while resolving detail to 20 metres, or down to 1 metre over a smaller area. Current Earth-observation efforts include NASA's Terra and Aqua satellites, the China–Brazil Earth Resources Satellite programme, the Disaster Monitoring Constellation (DMC) and France's SPOT satellite.

Radar systems, unlike optical systems, can penetrate cloud cover, which is common in the tropics. The microwave signals transmitted by Synthetic Aperture Radar (SAR) satellites, such as TerraSAR-X, are reflected back by leaves, branches and trunks to the satellite's receivers. Such satellites have historically been expensive, costing

between £250 million (US\$384 million) and £500 million each in orbit. Fortunately, a new generation of lower-cost radar mini-satellites is due to be launched within the next year. These include the British NovaSAR-S range (£45 million launched and insured). The European Sentinel-1 mission and the Japanese ALOS-2 satellite each cost about ten times this.

Radar and optical systems are both needed. An early warning constellation of five tropical-orbiting radar satellites should provide daily scanning of tropical forests at 5–20-metre resolution, irrespective of weather, to monitor logging in real time. We estimate that this could be done for £200 million, plus a dedicated ground crew. Ideally, these data should be managed by an international consortium that includes developed and developing countries, and would be run under the auspices of the UNFCCC.

Low-resolution optical satellites can be constructed cheaply to monitor forests weekly or monthly throughout the year⁷. Existing satellite systems, including NASA's MODIS, DMC, SPOT or Landsat, could contribute, with much of the data made available for free.

But to provide regular ground data with which to calibrate all these satellite observations and reduce uncertainties in their carbon estimates, major efforts will also be required by those countries participating in REDD+.

As discussions continue this year, we urge policy-makers to back the right satellites and strategies to monitor and save the world's forests. ■

Jim Lynch is professor at the Centre for Environmental Strategy, University of Surrey, Guildford, UK. **Mark Maslin** is professor in the Department of Geography, University College London, UK. **Heiko Balzter** is director of the Centre for Landscape and Climate Research, University of Leicester, UK. **Martin Sweeting** is director of the Surrey Space Centre, University of Surrey, Guildford, UK. e-mail: j.lynch@surrey.ac.uk

1. Friedlingstein, P. *et al.* *Nature Geosci.* **3**, 811–812 (2010).
2. Nellemann, C., INTERPOL Environmental Crime Programme (eds). *Green Carbon, Black Trade* (United Nations Environment Programme, GRID-Arendal; 2012).
3. Gullison, R. E. *et al.* *Science* **316**, 985–986 (2007).
4. Canadell, J. G. & Raupach, M. R. *Science* **320**, 1456–1457 (2008).
5. Global Observation of Forest and Land Cover Dynamics. *Reducing Greenhouse Gas Emissions from Deforestation and Degradation in Developing Countries* (GFOC-GOLD, 2012); available at go.nature.com/fk3ixp.
6. Herold, M. *et al.* *Carbon Balance Mgmt* **6**, 13 (2011).
7. Maslin, M. & Scott, J. *Nature* **475**, 445–447 (2011).

The authors declare competing financial interests: for details see go.nature.com/x6dtxy.



Heed the father of cladistics

The way Willi Hennig discovered evolutionary relationships should not be forgotten, say **Quentin Wheeler, Leandro Assis and Olivier Rieppel**.

In *On the Origin of Species*, Charles Darwin proclaimed that “our classifications will come to be, as far as they can be so made, genealogies”. That turned out to be easier said than done. Even as late as the 1970s, biologists were still grouping animals and plants largely on the basis of overall physical similarity and whether they possessed or lacked certain traits, such as a backbone or the ability to produce flowers.

The German entomologist and palaeontologist Willi Hennig transformed the classification of organisms into the rigorous science of cladistics^{1–5}. His book *Phylogenetic Systematics*⁶, published in 1966, laid out how to construct phylogenetic trees and how to use their branching patterns as the basis for classifications.

Paired with DNA sequencing, Hennig’s theories revolutionized our understanding of the relationships among the nearly two million species known today. In the history

of biological classification, the little-known Hennig arguably deserves a place alongside Aristotle, Carl Linnaeus and Darwin.

But two key messages from his book have been lost in the nearly half-century since it was published: the importance of detailed studies of the development and evolution of complex characters, such as the horn of a rhinoceros or the pincer of a fiddler crab; and the use of all relevant evidence — molecular, anatomical, fossil and developmental — in mapping evolutionary relationships. Too often these days, DNA information is favoured over everything else, and when conflicts arise between DNA-based analyses and those reliant on morphology, the former is frequently assumed to be correct, even though many uncertainties surround the molecular basis of evolution.

Hennig was born 100 years ago this week, on 20 April 1913. In celebration of his impact on phylogenetics and classification, we urge

biologists to heed his call to embrace all the relevant data.

BIOLOGICAL BEGINNINGS

Hennig, who died in 1976, was born in the village of Dürrhennersdorf in Germany⁷. He worked as a volunteer at the Dresden Museum, where he became fascinated by the extraordinary diversity of flies, and later as a researcher at the German Entomological Institute in Berlin until he was conscripted when the Second World War broke out in 1939. Badly injured on the Russian front, Hennig ended up serving in the German military medical services after his recovery.

Germany surrendered in May 1945 and Hennig, along with other German soldiers in Italy, came under Allied control. Recognizing his entomological expertise, the British assigned him to their malaria research unit, where he remained until his release in October of that year. While in this unit, Hennig produced a handwritten draft of his 1950 work, *Grundzüge*, the forerunner of his 1966 book, based in part on his observations of insect evolution.

From 1947, Hennig commuted for an hour-and-a-half each way from his home in West Berlin to the German Entomological Institute in East Berlin. The erection of the Berlin Wall in 1961 halted his commute, as well as plans to make him director of the institute. In 1963, he became head of a new department for phylogenetic research at the State Museum of Natural History in Stuttgart.

Hennig received numerous awards in his lifetime for succeeding where generations of natural historians had failed, including gold medals from the Linnean Society of London and the American Museum of Natural History. Yet the remarkable achievements of this quiet, unassuming man are mainly appreciated only within systematic biology. That cladistics is widely used today is thanks in part to Hennig’s American ‘bulldogs’, who promoted his ideas: Norman Platnick, a spider expert, Gareth Nelson, a fish expert, and James Farris, a theorist in phylogenetics¹.

Before the 1970s, many biologists viewed phylogenetic trees and classifications with suspicion, and for good reason. The construction of trees was often based more on a hunch than on testable hypotheses, and classifications did not strictly reflect branching patterns. The class Reptilia, for example, included crocodiles but excluded birds, even though the anatomical evidence strongly suggested that crocodiles and birds share a common ancestor. Meanwhile, all that held the ‘invertebrates’ together was their lack of a backbone. Like Darwin, Hennig believed that groupings of species could and should be strictly genealogical. He reasoned that such phylogenetic classifications could have the same organizing function in biology ►

► that the periodic table of elements had been in chemistry.

This led him to define a 'monophyletic group' as one that included an ancestral species and all (and only) its descendant species. He also insisted that all the members of a monophyletic group share at least one evolutionary novelty that arose in the common ancestor, such as wings in insects. Until this point, monophyly loosely meant sharing a common ancestor, whether or not all the descendants were included. Using Hennig's approach, Reptilia could be considered a monophyletic group only if it included birds. At the time, his proposal was severely criticized by several prominent biologists, including Ernst Mayr, yet this use of monophyly is now standard practice.

A great challenge for systematic biologists is to work out where in a genealogy an evolutionary novelty first appears and at what level on the phylogenetic tree it is informative. The fact that an organism has six legs, for instance, indicates that it belongs to the group Hexapoda (insects and their near relatives), but says nothing about whether it is a fly (order: Diptera). Hennig emphasized that every evolutionary novelty that arose in a common ancestor, either in its original or modified form, is potentially informative at the appropriate phylogenetic level, and advocated the use of all available data — from biochemical pathways to skeletal structures and genetically controlled behaviours. In fact, he argued that living species provide a far richer source of historical information

than fossils, in which only a small fraction of features are preserved.

HARD TRUTH

Hennig never implied that fossil evidence should be ignored. But his take on fossils surprised many contemporary palaeontologists and evolutionary biologists. In the early twentieth century, and to a lesser extent, even while Hennig was writing his book, the fossil record was treated as if it revealed the truth of evolutionary history. We think that — as Nelson⁸ argued almost a decade ago — DNA has similarly come to be wrongly viewed as the key to phylogeny, rather than just one among several sources of evidence.

DNA sequencing has clear advantages: vast amounts of data can be collected quickly and cheaply, and sequences seem to offer a more objective measure than assessments of complex morphological features. But DNA analyses also involve subjective judgements. For instance, the assumptions used to produce models of molecular evolution, on which such analyses are based, may differ. And when conflicts arise between DNA analyses and those based on morphology, fossils or ontogeny, there is no theoretical justification for favouring one source of data over another.

Take the relationships among the

"DNA has come to be wrongly viewed as the key to phylogeny, rather than just one among several sources of evidence."

some 9,000 scaly reptile species. Iguanas, chameleons and their close relatives had long been placed in the lower branches of the phylogenetic tree, but various DNA studies conducted over the past decade have suggested that they are actually higher up, near snakes and Gila monsters (a venomous lizard species). A recent analysis of more than 600 morphological characters now indicates that the traditional placement of iguanas in the tree is correct after all⁹. The original phylogeny certainly seems more plausible; for the DNA-based branching pattern to be correct, an extraordinarily large number of complex forms would have had to have independently evolved back into an ancestral state.

Furthermore, used in isolation, DNA-sequencing technologies cannot fully explain the characteristics, history and origins of complex characters, such as xylem vessels, seeds, flowers or feathers, that have arisen over millions of years. It is hard to see how biologists can understand the many ways in which organisms have adapted to their environments without analysing their physical adaptations.

The progress in exploring and classifying biodiversity that Hennig's work unleashed is just the beginning: biologists estimate that eight million to ten million plant and animal species remain unknown to science, and possibly even more microbes¹⁰. To understand the details of evolutionary history that help to explain this incredible diversity, biologists should remember Hennig's message: the clues are in evolutionary innovations at all levels, molecular, anatomical, developmental and behavioural. ■

Quentin Wheeler is professor at the Schools of Sustainability and Life Sciences, Arizona State University, Tempe, Arizona, USA.

Leandro Assis is professor at the Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. **Olivier Rieppel** is curator of evolutionary biology at the Field Museum of Natural History, Chicago, Illinois, USA.

e-mail: quentin.wheeler@asu.edu

1. Hull, D. L. *Science as a Process* (Univ. Chicago Press, 1988).
2. Nelson, G. & Platnick, N. *Systematics and Biogeography* (Columbia Univ. Press, 1981).
3. Schuh, R. T. & Brower, A. V. Z. *Biological Systematics* (Cornell Univ. Press, 2009).
4. Rieppel, O. C. *Fundamentals of Comparative Biology* (Birkhäuser, 1988).
5. Wheeler, W. C. *Systematics* (Wiley Blackwell, 2012).
6. Hennig, W. *Phylogenetic Systematics* (Univ. Illinois Press, 1966).
7. Schmitt, M. *From Taxonomy to Phylogenetics: Life and Work of Willi Hennig* (Brill, 2013).
8. Nelson, G. in *Milestones in Systematics* (eds Williams, D. M. & Forey, P. L.) 127–147 (CRC Press, 2004).
9. Gauthier, J. A. et al. *Bull. Peabody Mus. Nat. Hist.* **53**, 3–308 (2012).
10. Wheeler, Q. D. et al. *Syst. Biodivers.* **10**, 1–20 (2012).



Willi Hennig stressed that every evolutionary novelty, from a spider's web-making ability or a giraffe's neck to a fiddler crab's pincer, is potentially informative in mapping evolutionary relationships.





E. O. Wilson studying ant behaviour in 1975.

CAREERS

A guide to the life scientific

Stuart Pimm applauds eminent biologist E. O. Wilson's pragmatic and passionate career advice.

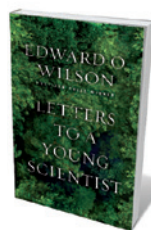
My mind races through possible conversations as I see my phone call is from E. O. Wilson. Pulitzer Prize winner, author of *Sociobiology*, laureate of the top international prizes for environmental sciences, widely acknowledged as one of the great scientists of our age, emeritus professor still mounting incisive attacks on those who challenge his ideas on evolution: what could he want to discuss with me? "Ants" I told myself quickly, organisms about which I know nothing. He did not disappoint, grilling me about my experiences in the Hawaiian islands, where ecosystems assembled without ants.

I expected *Letters to a Young Scientist* to be about ants, too. It is. Ants appear every few pages in its 21 short chapters. Advice flows as Wilson identifies them, hunts down rare and evolutionarily special ones from Sri Lanka to the Australian outback, and makes his lab smell by putting chemicals on their bodies that trigger their ejection from nests as dead, even when they are not. They are the focus of global explorations and innumerable simple experiments.

In this fund of practical and philosophical guidance distilled from seven decades of experience, Wilson provides exactly the right mentoring for scientists of all disciplines — and all ages. This is a very personal story, from boyhood Scout camp in Alabama to Harvard University, replete with autobiographical

details, stories of the famous, the eccentric (and usually both), of success and, far more often, of failure and frustration.

Wilson takes on that "great bugbear for many would-be scientists" — mathematics — in Chapter 2. He grants that some disciplines require exceptional fluency, but warns that a strong mathematical background does not guarantee success. He conjures a field of dreams. Build engaging ideas and the mathematicians will come. "For



Letters to a Young Scientist

EDWARD O. WILSON
Liveright/W. W.
Norton: 2013. 242 pp.
\$21.95/£15.99

every scientist ... there exists a discipline of science for which that level of mathematical competence is enough to achieve excellence," Wilson posits as a principle.

Nor is exceptional IQ essential. He writes: "let me suggest ... that many of the IQ brightest join ... MENSA and work as auditors and tax consultants." Yes, it helps to be smart, but one can have too much of a good thing. Wilson thinks that those who are too smart don't have to sweat to do science as students, something that he and I agree is necessary to build the requisite character.

There are admonitions for later stages in careers. "Avoid department-level administration." "Make excuses, dodge, plead, trade." "Spend extra time with students who show talent and interest in your field of research." "Consider carefully job offers that include fewer ... administrative responsibilities." This book will quickly be on the *Index Librorum Prohibitorum* of college deans.

Wilson sees daydreaming as a key exercise for success, and suggests: "Make talking to yourself silently a relaxing pastime." My personal experience suggests that silence is not an essential ingredient of that recipe, although my wife would appreciate it if it were. He also urges his putative young scientist to explore. He won the coveted Explorers Club Medal, despite never venturing to the poles or making first contact with an Amazonian tribe. But his exploring for biodiversity has made many aware that scientific frontiers are all around us.

Can creativity be engineered? Wilson visits the Santa Fe Institute in New Mexico, where personal office space is deemed distasteful, and Google's headquarters in Mountain View, California. When I visited, I tried not to overindulge in *ad libitum* soft drinks, coffee and nuts. And I still haven't got over the advice on writing better computer code pasted above the men's urinal. These places impress Wilson too. Nonetheless, he believes that the creative process "arises and for a while generates in a solitary brain".

"You will make mistakes. Try not to make big ones," he concludes. Amen to that. Yet Wilson begins the book with the most important advice of all: "enduring passion will never fail you." It hasn't failed him.

This is no pompous, deeply philosophical treatise on how great ideas develop. Wilson shares his simple love for ants and their natural history, revelling in them without hesitation. Everything else follows. ■

Stuart Pimm is professor of conservation at the Nicholas School of the Environment, Duke University, Durham, North Carolina, USA, and author of *The World According to Pimm: a Scientist Audits the Earth*. e-mail: stuartpimm@me.com



The pituri plant has long been known for its narcotic properties.

MEDICAL HISTORY

Feeling no pain

John Carmody enjoys an exhibition that charts the trajectory of anaesthesia from its botanical beginnings.

“Publish or perish” is more than a crass modern summary of academic career-building. Without the written or printed word, work can be forgotten for centuries — or even forever.

This message is beautifully underlined in *History of Anaesthesia*, an exhibition on pain relief at the University of Sydney, Australia, that draws on the institution’s splendid Rare Books and Special Collections, as well as the equipment museum maintained by the Australian Society of Anaesthetists. Through some 200 books, official reports, quaint and serious apparatus, and illustrations spanning more than five centuries, the show explores the age-old quest to dull agony.

➔ **NATURE.COM**
For more on the science of pain, see: go.nature.com/np9pcy

The exhibits are as much a reminder of this great aspiration

as of the panoply of methods — exotic, futile and pragmatic — designed to fulfil it. But it is the books that best reveal the thinking of our intellectual forebears.

Because almost all that is ‘physic’ is derived from plants — the pain-relieving potions concocted from, say, the opium poppy — the exhibition begins with medical botany as systematized in early ‘herbals’. Perhaps the most precious is a book translated from Latin by the English surgeon John Halle. *The herbal: an expositive worke after the order of the alphabet*, published in 1565, is actually an extract from Lanfranco of Milan’s seminal thirteenth-century surgical treatise *Chirurgia parva*.

Halle, an Elizabethan progressive,

History of Anaesthesia

Fisher Library,
University of Sydney,
Australia.
Until 12 May 2013

championed sourcebooks in the vernacular to promote learning. An important figure as modern medicine began to emerge in England, he summarized his aim as “the edification and building up of good science, and to the subversion of all haters and abusers of the same”.

Among other traditions of botanical medicine, an intriguing Australian example is the little paper “Pituri and Duboisia”, which the parasitologist and surgeon Joseph Bancroft read to the Queensland Philosophical Society in 1877. Bancroft refers to the pituri plant (*Duboisia hopwoodii*), widely used by indigenous Central Australians as “a stimulating narcotic”.

The renowned German botanist Ferdinand von Mueller, who settled in Melbourne, identified Bancroft’s samples and surmised that *Duboisia*’s properties might be similar to the narcotic effect of stramonium, an alkaloid extracted from a plant of the nightshade family and used for centuries to relieve asthma and pain. We learn from the paper that Bancroft, testing it on his domestic pets, found the effects “strange”, adding that “they seem blind, or nearly so, with a widely-dilated pupil”. Afterwards, while dosing several of his ophthalmic cases, he “found an action of great rapidity” and quickly alerted receptive colleagues elsewhere in Australia. The country’s plantations of *Duboisia* species and hybrids still supply around 70% of the global stocks of the alkaloids scopolamine and atropine.

As chemistry developed, the active constituents were purified — crucially for anaesthesia, which was perhaps the most pharmacologically dependent medical speciality. Equally important were fundamental discoveries about atmospheric gases and respiration — both crucial to inhalational anaesthesia. Here, the German-Swedish chemist Carl Wilhelm Scheele’s name should be pre-eminent: a 1780 English translation of his *Chemical Observations and Experiments on Air and Fire* is rightly part of the exhibition. He would be better remembered had he not been usurped by Antoine Lavoisier, whom he told how to prepare oxygen (in a letter made public in Paris in 1992).

Beyond revelling in the beauty and nostalgia of remarkable old books, this exhibition reminds us that perhaps no area of medicine relies more than anaesthesia on a diversity of scientific understanding. The modern management of pain — the “perfect misery, the worst of evils”, as John Milton described it in *Paradise Lost* — has been an intellectual venture. ■

John Carmody is an honorary associate professor in the disciplines of anatomy and physiology at the School of Medical Sciences of the University of Sydney, Australia.
e-mail: john.carmody@sydney.edu.au

PUBLISHING

The word on popular science

To mark UNESCO's World Book and Copyright Day on 23 April, **Josie Glausiusz** asks science editors at leading book publishers about trends and technology.

Animated, interactive, and evolving at the reader's command: this is the popular science book of the near future, according to T. J. Kelleher, the science commissioning editor at Basic Books in New York.

"We can expect to see new kinds of books that would have been impossible on paper," Kelleher says. "The straightforward book that is nothing but words is not going any place; people obviously like to read, but there are other ways they can experience information." So a mathematics e-book might incorporate a user interface that lets readers change variables: for example, the density, volume and speed of an asteroid hurtling towards Earth.

Basic Books has already taken the plunge. In October 2012, the company published a version of Richard Feynman's *Six Easy Pieces* for the iPad that incorporates audio recordings of Feynman's physics lectures at the California Institute of Technology in Pasadena — and 5,000 photographs of the revered Nobel laureate delivering them.

My poll of science editors at major publishers revealed enthusiasm for apps and e-books. In September 2011, Farrar, Straus and Giroux (FSG), in a joint imprint with *Scientific American*, launched its first app. *Journey to the Exoplanets* has more than

100 illustrations, as well as essays and audio clips from planetary scientists, including Caleb Scharf and Lawrence Krauss.

The electronic format, says Latha Menon, a senior commissioning editor at Oxford University Press in Oxford, UK, is "ideal for transmitting science ideas", because animations can be used to illustrate processes such as natural selection. In one example, she says, "you can actually see how the light-grey moths are picked off" in preference to dark moths camouflaged by soot-grimed tree trunks, an adaptation seen during the industrial revolution.

And, notes FSG senior editor Amanda Moon, "the thirst for knowledge about topics that fall under the broad umbrella of 'science' is intensifying. We all have questions about evolution, technology, psychology, animal behaviour." Others echo her, although with none prepared to divulge exact sales figures, it is hard to know whether their enthusiasm reflects profits.

The science books that find a devoted readership these days, says Moon, "boast provocative arguments grounded in original research, are character-driven narratives and, in many cases, offer counterintuitive angles on seemingly familiar topics". She

cites Kevin Dutton's *The Wisdom of Psychopaths* (Scientific American, 2012), which suggests how we can learn from "functional psychopaths" such as some neurosurgeons and lawyers. Authors, she adds, now have many more opportunities to get the word out about groundbreaking ideas and research through platforms such as Facebook, Twitter and Goodreads.

Many of today's readers are prepared for a challenge. "What's interesting," says Menon, "is that there seems to be a significant market of people out there who perhaps studied some science or even did a science degree", and then went off to do something else. That readership, Menon adds, is sophisticated, demanding books that are "conceptually, a little stretching": for example, Jim Baggott's *The Quantum Story* (Oxford Univ. Press, 2011), a history of quantum theory from 1900 to the present in 40 key moments. Epigenetics, neuroscience and particle physics — the last prompted by the discovery last year of a Higgs boson — are currently big, she says. Although, as in every other field, science book publishing is dominated by television tie-ins and celebrity authors such as Brian Cox and Stephen Hawking.

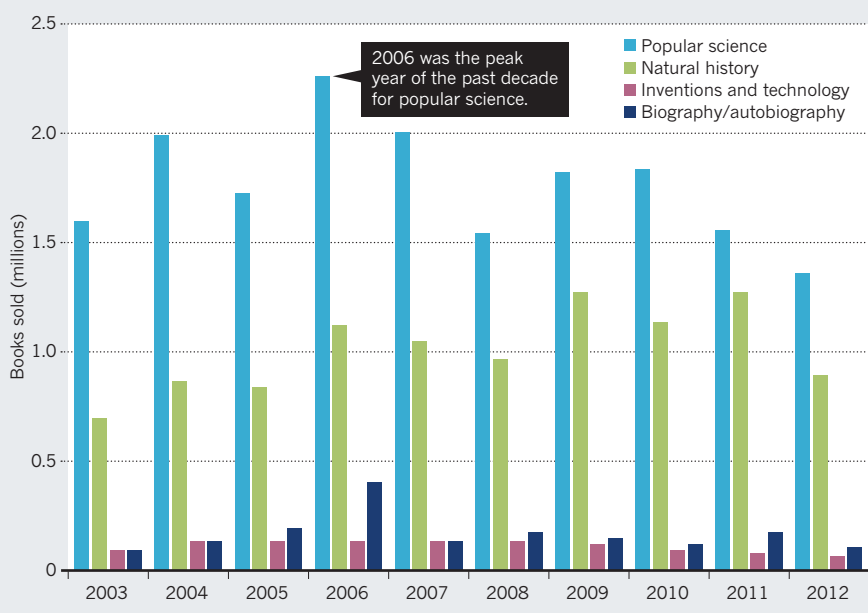
Christie Henry, editorial director of sciences and social sciences at the University of Chicago Press in Illinois, concurs that, at present, the public appetite for science seems voracious (see 'Science book sales'). "The most pressing issues we face as a global society rely on science to some extent," she says. Books on water rights, infectious diseases, conservation and marine science are all proving popular, she explains.

All agree that well-written science books are unlikely to disappear, even in an age in which the average person's eyes seem to be perpetually locked on mobile devices. "The quality of the writing and the quality of the narrative matter," says Menon. Classics such as Richard Dawkins' *The Selfish Gene* (Oxford Univ. Press, 1976) and Primo Levi's 1975 *The Periodic Table* continue to sell well. They have, says Menon, "a literary quality, a structure and a beauty. There are just a few scientists who can write like that, and that is something to be treasured and valued. I would hope that that continues to have a home." ■

Josie Glausiusz is a writer based in Israel.
e-mail: josiegz@gmail.com

SCIENCE BOOK SALES

Figures from the United Kingdom and Australia illustrate the continued healthy market for general-interest science books.



SOURCE: DAVID WALTER/NIELSEN BOOK SCAN

Correspondence

Overfishing in west Africa by EU vessels

China is not the only culprit threatening the marine ecosystems of west Africa through overfishing (*Nature* **496**, 18; 2013). The region has been the fish basket of southern Europe since the 1960s, and the European Union (EU) is going to great lengths to renew its fishery agreements with African nations.

These agreements are questionable because they promote the export of African marine resources into the EU to the detriment of local economies and food sovereignty. Morocco has allowed EU vessels into the national waters of neighbouring Western Sahara, effectively hijacking its fish stocks.

The uncontrolled harvesting of pelagic fish and octopus by hundreds of foreign fishing vessels is having a deep ecological impact on the Atlantic Ocean's Canary Current upwelling marine ecosystem, a major biodiversity hotspot. It is also destroying vast amounts of by-catch.

Research programmes are under way to quantify this collateral damage. Some highly vulnerable and endangered marine vertebrates are affected, including monk seals, several shark species, dolphins and sea turtles, and hundreds of thousands of Macaronesian and Palaearctic seabirds that breed and overwinter along the west African coast (see, for example, J. Zeeberg *et al.* *Fish. Research* **78**, 186–195; 2006).

We therefore urge EU parliament members to rethink their African fishing policies, properly taking into account the devastating political, social and ecological impact of their large industrial fisheries.

Raül Ramos *Centre for Functional and Evolutionary Ecology (CEFE), CNRS, UMR5175, Montpellier, France.*
raul.amos@cefe.cnrs.fr

David Grémillet *CEFE–CNRS Montpellier, France; and PFI/O, DST/NRF Centre of Excellence,*

University of Cape Town, South Africa.

Scale of global road map is impractical

William Laurance and Andrew Balmford's proposed global zoning map (*Nature* **495**, 308–309; 2013) for road planning would be too crude in scale to be useful for practical, on-the-ground decisions.

Positive outcomes for biodiversity from road building are more likely to be achieved in a local context. Decisions about road developments are not made on a global scale, and rarely even on a national level.

Any road planned with environmental and social considerations in mind is better than one designed without them. But decision-makers should not shift their focus too far from the major problems that proposed roads might create.

Malgorzata Blicharska *Swedish University of Agricultural Sciences, Uppsala, Sweden.*
malgorzata.blicharska@slu.se

Uphold standards for lab reagents

Scientists assume that reagents supplied commercially have been rigorously tested for purity. We have found that this assumption can be risky.

Our laboratory purchased 3 milligrams of a biomolecule from a reputable supplier to investigate its effect on nerve cells. Anomalous results forced us to test its purity. Using mass spectrometry and qualitative western blotting, we discovered unacceptable levels of contamination with several other biologically active molecules.

Several research groups have published important mechanistic conclusions based on the activity of this molecule. Contamination with other active components would call these results into question.

Reagents that are misleadingly labelled waste researchers' funds and time. Suppliers should declare the accurate composition and modes of analysis for every batch of their products.

Sally Roberts, Heidi Fuller *Keele University, Staffordshire, UK; and Robert Jones & Agnes Hunt Orthopaedic Hospital, Oswestry, UK.*
sally.roberts@rjah.nhs.uk
Bruce Caterson *Cardiff University, UK.*

Crowd-sourced soil data for Europe

Since its launch by the UK Natural Environment Research Council in June 2012, the free iPhone app *mySoil* has attracted more than 2.6 million web hits and 12,500 dedicated users (see go.nature.com/ty1ukf). Next month, a new version will extend its coverage to soils in the European Union (EU).

Users of *mySoil* can currently view soil maps of the United Kingdom (excluding Northern Ireland) that provide regional information on soil depth, texture, pH and organic-matter content, and on vegetation habitats. They can upload photos and descriptions of their local soils. More than 500 entries have so far contributed to this valuable data bank of soil properties in different localities.

The remarkable success of collecting data in this way reflects widespread public interest and promotes communication with the science community.

With infrastructure covering an estimated average of 9% of the EU — often spread across the most productive soils — and with sustainable land management becoming increasingly urgent, *mySoil's* democratization of soil data is a timely development.

Wayne Shelley, Russell Lawley *British Geological Survey, Nottingham, UK.*
David A. Robinson *Centre for Ecology & Hydrology, Bangor, UK.*
davi2@ceh.ac.uk

Chinese institute clarifies patent case

As director and deputy director of the Institute of Microbiology, Chinese Academy of Sciences (IMCAS), we would like to clarify some points in your discussion of our involvement in an intellectual-property case (*Nature* **492**, 323; 2012).

Without our knowledge, Shandong Hilead Biotechnology in Laiyang — a recipient of the dodecanedioic acid technology invented by IMCAS, with which Hilead collaborated to further improve this technology — filed a patent application in 2010 that listed both of us and five others as inventors (patent CN201010160266.4).

When we discovered this, we demanded that our names be dropped from the application because we had not made any technical contribution to the patent. Hilead filed an application for a change of inventors listed on the patent on 25 January 2011 to the State Intellectual Property Office of China, who approved the change on 14 February 2011 (file 2011021000026950). So the removal of our names occurred a few months before two employees at Cathay Industrial Biotech in Shanghai filed a lawsuit over the patent in June 2011.

Hilead officially apologized on 17 August 2011 in a letter to IMCAS for using our names without consent. The court decision on the lawsuit exempted us from civil liability and no apology was required from us (Beijing Higher People's Court decision 3975, 16 November 2012).

Li Huang, Shuang-Jiang Liu *IMCAS, Beijing, China.*
huangli@im.ac.cn

CORRECTION

A declaration of competing financial interests was omitted from the Comment 'A jump-start for electroceuticals' (*Nature* **496**, 159–161; 2013). This has been added online.

ARCHAEOLOGY

A potted history of Japan

The discovery of lipids on ceramic fragments from the Japanese Jōmon period provides the earliest evidence for the use of pottery for cooking and may prompt a rethink of some aspects of human innovation. [SEE LETTER P.351](#)

SIMON KANER

The invention of pottery is one of the great technological innovations in human history. The recognition that fire could be used to transform soft clay into impermeable ceramics represented a cognitive shift in the way people engaged with the material world around them¹. Until recently, archaeologists usually regarded pottery as being associated with agriculture. Pottery usage among hunter-gatherers was considered somewhat anomalous and counter-intuitive; fragile pots did not seem to have a place in the mobile lifestyles thought to characterize most human existence before the advent of farming villages during the Neolithic, from about 10,000 years ago in the eastern Mediterranean. But the discovery of lipids on ceramic vessels in East Asia dating from the Late Pleistocene, about 15,000–12,000 years ago, reported by Craig *et al.*² on page 351 of this issue, suggests that some hunter-gatherers used pots for cooking. The report also provides a demonstration of how science should be integral to our piecing together of history*.

Craig and colleagues report the first recovery and analysis of lipids from charred deposits — the remains of prehistoric meals — stuck on the surfaces of potsherds of some of these ancient pottery containers. The authors' description of the lipids, supported by stable-isotope analysis, indicates that Late Pleistocene peoples in what is now the Japanese archipelago were using pottery to cook fish, both freshwater and marine. With claims of pottery in East Asia extending back even further than this, perhaps as far as 20,000 years ago³, the application of rigorous scientific analyses to these materials is of great relevance to furthering our understanding of the cultural contexts in which pottery was first made and used.

Indications of Late Pleistocene pottery are not completely new, but little is known about what these vessels were used for. In the late 1960s, thin ceramic potsherds associated with tiny obsidian stone blades, or microliths, were excavated at the Fukui Cave in southwestern Japan, dated to around 12,000 years ago⁴.

*This article and the paper under discussion were published online on 10 April 2013.



Figure 1 | Fish soup from the Late Pleistocene. Ceramic pots dating to the Late Pleistocene have been found in several sites throughout East Asia. This Incipient Jōmon pot from Kubodera-minami in Niigata prefecture in Japan is around 15,000 years old. Craig and colleagues' analysis² of such pottery has revealed lipid deposits that suggest the pots were used to cook fish.

But it was the application⁵ of a new type of radiocarbon dating using accelerator mass spectrometry — made famous by its use in the reassessment of the date of the Turin Shroud — to carbonized accretions on the surface of potsherds from a small undecorated vessel from northern Japan, at Odai Yamamoto, that set the debate over the existence of Late Pleistocene pottery alight.

When the early dates for pottery use at Fukui Cave and other sites were first announced, many Japanese archaeologists were reluctant to accept them. At the time, chronologies of Japanese prehistory were based largely on the recognition of different styles of pottery. This framework had its origins in pre-war research that avoided explicit chronologies so as not to clash with the version of Japanese history laid down in two eighth-century texts, the *Kojiki*

and *Nihon Shoki*. Under the militaristic regime of the 1930s and '40s, questioning these sacred sources was unacceptable. Even after 1945, as in Europe, authority figures found it impossible to reconcile new radiocarbon-based dating of archaeological finds with their own style-based chronologies^{6,7}.

Furthermore, until the 1990s, the ancestry of most of the current population of the Japanese archipelago was thought to trace back to the rice-farming peoples of the Yayoi period, then dated from about 300 BC to about AD 300. The arrival of paddy-rice agriculture in the first millennium BC was regarded as marking a major break in population and culture from the preceding Jōmon period, named after the eponymous cord-marked pottery from the time (Fig. 1). The Jōmon people were largely considered to be primitive aboriginals who had little to do with the development of later Japanese culture⁸.

But the plethora of new archaeological discoveries associated with the post-war development boom changed the image of Jōmon society. The discovery and excavation in the early 1990s of one of the largest known Jōmon settlements, at Sannai Maruyama, just a few kilometres from Odai Yamamoto, produced the remains of buildings and burials, and evidence for spatial planning and a rich material culture, including tonnes of potsherds. Radiocarbon dating demonstrated that the site was continuously occupied, at varying levels of intensity, for nearly two millennia⁹. These findings were announced as Japan was coming to terms with the end of a seemingly endless trajectory of economic growth, and the Jōmon culture captured the public imagination as offering an alternative vision of sustainable community life in the Japanese archipelago. The Jōmon became embraced as ancestral and in tune with the natural environment of Japan¹⁰.

Soon after, however, Japanese archaeology went into crisis. In 2000, claims made in the 1980s and '90s of increasingly early dates for 'Early Palaeolithic' sites across eastern Honshu were shown to be the result of fraudulent activity (for a summary, see ref. 11). The revelations sent shock waves through the field and undermined public trust in the archaeologists who had worked so hard to

create a new past for Japan since 1945, when the version of history provided by the *Kojiki* and *Nihon Shoki* was no longer accepted as defining early Japanese history¹².

Specialists in the techniques that are used to identify fatty acids on archaeological finds were implicated in this fraud, bringing the reliability of these methods into question. The unfortunate situation led to an enhanced recognition that archaeologists need the skills to critically assess scientific information, and that this must be open to international scrutiny. The analyses presented by Craig *et al.* not only restore faith in the use of lipid-analysis techniques for making inferences about archaeological objects, but also are representative of the tremendous resource offered by Japanese archaeology for understanding the human past.

Our knowledge that ceramic containers were being made and used by hunter-gatherers in the Late Pleistocene in various parts of East Asia — from Japan to far eastern Russia and north and south China — means that pottery usage among hunter-gatherers is no longer seen as anomalous in the Old World. In fact, there may be evidence for routes of the introduction of pottery into Europe that are not associated with the introduction of farming¹³.

Yet many questions remain about how and why pottery was invented. For example, the earliest known ceramics are not containers at all, but fragments of figurines, such as those dating to 29,000 years ago from the Dolní Věstonice site in the Czech Republic¹⁴. Although the earliest known containers were probably used for cooking, the special significance given to this new technology from the outset should not be overlooked. The fact that the prehistoric foragers who were using these implements were not rigorous in their washing up also allows us, to some extent, to analyse the contents of their meals, as Craig and colleagues have begun to do. We now need to extend such evaluation to the cultural meanings of what they were eating in order to further understand why they needed pottery containers¹⁵. ■

Simon Kaner is at the Centre for Archaeology and Heritage, Sainsbury Institute for the Study of Japanese Arts and Cultures, and at the Centre for Japanese Studies, University of East Anglia, Norwich NR4 7TJ, UK.
e-mail: s.kaner@uea.ac.uk

1. Kobayashi, T. *Jōmon Reflections: Forager Life and Culture in the Prehistoric Japanese Archipelago* (eds Kaner, S. & Nakamura, O.) (Oxbow, 2004).
2. Craig, O. E. *et al.* *Nature* **496**, 351–354 (2013).
3. Wu, X. *et al.* *Science* **336**, 1696–1700 (2012).
4. Aikens, C. M. & Higuchi, T. in *Prehistory of Japan* 99–104 (Academic, 1982).
5. Odai Yamamoto I Site Excavation Team (eds) *Archaeological Research at the Odai Yamamoto I Site* (Kokugakuin Univ., Tokyo, 1999).
6. Barnes, G. L. *Antiquity* **64**, 929–940 (1990).
7. Renfrew, C. *Before Civilisation* (Penguin, 1973).
8. Morse, E. S. *Traces of an early race in Japan*. *Popular Sci. Mon.* **14**, 257–266 (1879).
9. Habu, J. *Antiquity* **82**, 571–584 (2008).
10. Hudson, M. J. in *Hunter-gatherers of the North*

Pacific Rim (eds Habu, J., Savelle, J. M., Koyama, S. & Hongo, H.) 263–274 (Senri Ethnol. Studies No. 63) (2003).

11. *Inspection of the Early and Middle Palaeolithic Problem in Japan* (Japan. Archaeol. Assoc., 2003).
12. Kaner, S. *Before Farming* **2**, 4 (2002).
13. Jordan, P. & Zvelebil, M. (eds) *Ceramics Before*

Farming: The Dispersal of Pottery Among Prehistoric Eurasian Hunter-gatherers (Left Coast, 2010).

14. Vandiver, P. B., Soffer, O., Klima, B. & Svoboda, J. *Science* **246**, 1002–1008 (1989).
15. Kaner, S. in *Ceramics Before Farming: The Dispersal of Pottery Among Prehistoric Eurasian Hunter-gatherers* (eds Jordan, P. & Zvelebil, M.) 93–120 (Left Coast, 2010).

ASTRONOMY

A cosmic growth spurt in an infant galaxy

One of the most extreme starburst galaxies in the early Universe has been identified and characterized. This system shows the rapid formation of a massive galaxy when the Universe was only 6% of its current age. [SEE LETTER P.329](#)

DESIKA NARAYANAN & CHRIS CARILLI

Starburst galaxies play a central part in our understanding of galaxy formation. These galaxies — which are defined by star-formation rates that exceed those of similarly massive normal galaxies by factors of 10 to 100 — are probably the cosmic ancestors of present-day massive elliptical galaxies, and the sites at which supermassive black holes grow most actively. Charting the evolution of these galaxies through space and time is a fundamental challenge for astronomers, and a task that is far from trivial. Simply finding these sources at early cosmic times requires very sensitive multiwavelength observations, clever use of data-selection techniques and a bit of luck. On page 329 of this issue, Riechers and colleagues¹ develop a technique to find massive starburst galaxies in the early Universe. They report the detection of the farthest known pure starburst galaxy, which existed a mere 880 million years after the Big Bang.

This discovery builds on a long history of observational efforts to find and characterize the physical properties of starbursts at high redshift (a measure of location in cosmic time). Since the original discovery of distant, dust-obscured starburst galaxies at high redshift in 1998, through the first sensitive infrared and submillimetre surveys of the sky^{2,3}, there has been an explosion of interest in the field. This has been motivated in part by the development of sensitive ground-based and space infrared telescopes, such as NASA's Spitzer Space Telescope, and the far-infrared and submillimetre capabilities of the European Space Agency's Herschel Space Observatory. Starburst galaxies are typically hidden at optical wavelengths, owing to copious amounts of obscuring dust; astronomers, therefore, make use of submillimetre and infrared telescopes that capture the wavelength regime in which most energy from these galaxies is emitted. During the past

decade, researchers have learned that these high-redshift starbursts reside in exceptionally massive haloes of dark matter (with masses equivalent to 10 trillion Suns), and that they correspond to the most luminous star-forming galaxies in the Universe.

Despite the marked advances in infrared and submillimetre-wave technology during the 2000s, measuring the exact distribution of redshifts of these galaxies has proved notoriously difficult. Older samples of these infrared- and submillimetre-wavelength-selected galaxies required heroic observational efforts to determine their redshifts: galaxies had to be identified in deep submillimetre surveys, and their precise position established with radio interferometers, before they could be observed with large optical telescopes to obtain a spectral line that could be used in redshift identification⁴. This process showed that high-redshift starbursts predominantly reside at a redshift of about 2 (roughly 3 billion years after the Big Bang). However, a nagging worry persists that radio selection has resulted in a low-redshift bias; that is, is there a substantial population of massive starbursts at earlier cosmic times that has been missed in recent studies?

Riechers and colleagues found starbursts at record-setting redshifts by using a clever technique that capitalizes on the fact that these systems are expected to be extremely dusty. As a part of the Herschel Multi-Tiered Extragalactic Survey (HerMES), the largest project assigned to the Herschel Space Observatory, the authors searched 21 square degrees of the sky, and looked for sources that were extremely red in the telescope's 250-, 350- and 500-micrometre-wavelength filters. The team (correctly) guessed that these were very early galaxies whose already red colours were accentuated by the expansion of the Universe. Moreover, the authors capitalized on new broadband (radio through submillimetre) spectrographs to obtain a redshift from

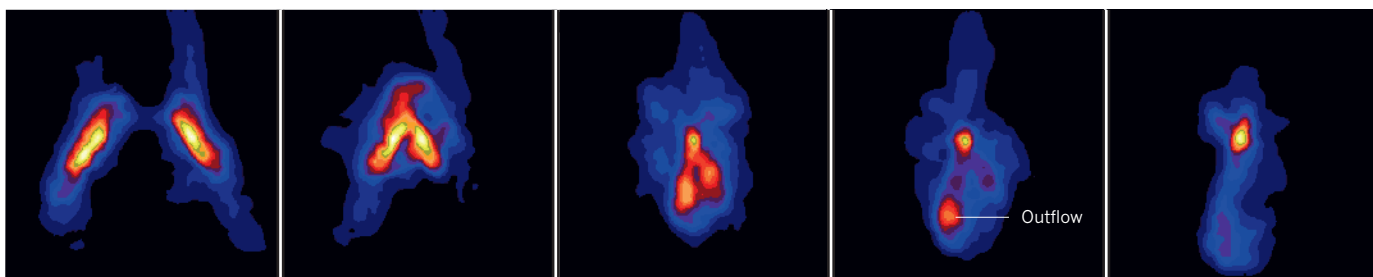


Figure 1 | A forming starburst. Mergers between two gas-rich galaxies can drive some of the most intense starburst galaxies in the Universe. This time sequence of images shows a numerical simulation¹⁰ of the expected view by the Atacama Large Millimeter/submillimeter Array in Chile of the star-forming molecular gas of a high-redshift galaxy merger that has physical

conditions comparable to those observed in the starburst galaxy detected by Riechers and colleagues¹. The simulated system drives a powerful molecular outflow from the nucleus after coalescence of the galaxies. The colours denote carbon monoxide emission intensity. Yellow corresponds to an intensity roughly 1,500 greater than black. Each panel is 50 kiloparsecs on a side.

atomic and molecular lines, revealing a source (monikered HFLS 3) at redshift 6.34 — the highest-redshift infrared-selected starburst galaxy found so far⁵.

A major added value of this work, however, is that the authors go well beyond simple ‘stamp-collecting’ goals of finding the earliest starburst galaxies, and obtain a huge suite of imaging and spectroscopic data from 12 international telescope facilities to characterize the physics that drives the evolution of the galaxy. With broadband imaging data that ranged in wavelength from the optical through to centimetre-wave photons, Riechers *et al.* derived an incredible star-formation rate for this system — about 3,000–5,000 solar masses per year. This contrasts sharply with the paltry roughly 2 solar masses per year of our own Milky Way. HFLS 3 may be the galaxy with the highest star-formation rate found so far.

Following on from this, the team analysed numerous emission lines from carbon monoxide (a standard tracer of star-forming gas in galaxies), an emission line from singly ionized carbon (one of the dominant coolants in galaxies) and emission lines from water. From an analysis of the excitation of carbon monoxide, as well as atomic-to-molecular line ratios, they determined that the gas participating in the compact starburst is probably illuminated by an intense radiation field, and may be thought of as a galaxy-wide star-formation region, comparable to the densest, most actively star-forming regions at the centres of giant molecular clouds in the Milky Way. In other words, this work has identified one of the most extreme star-forming environments in the Universe, at a time when the Universe was only 6% of its current age. The interstellar medium in HFLS 3 was already substantially enriched in the standard molecules that populate the interstellar medium in nearby spiral galaxies, such as our own. The authors further suggest that a merger between two gas-rich galaxies in the early Universe may be driving the starburst event, as well as a powerful molecular outflow (Fig. 1).

Perhaps the most exciting aspect of this and similar⁶ studies is the sheer potential in this field. Wide-band spectrometers on the Caltech Submillimeter Observatory in Hawaii,

the Plateau de Bure Interferometer in France, the Green Bank Telescope in West Virginia, the Karl G. Jansky Very Large Array in New Mexico and the Atacama Large Millimeter/submillimeter Array in Chile now allow for rapid calculation of redshifts through the detection of multiple molecular emission lines, while avoiding the redshift biases that plagued previous studies^{6,7}. Thus, capabilities will only continue to improve in the coming years with the Large Millimeter Telescope in Mexico and Cornell Caltech Atacama Telescope in Chile. Surveys should soon reveal the cosmic number-density evolution of high-redshift, optically obscured starbursts — a quantity that will place strict constraints on models of galaxy formation and evolution⁸. Furthermore, the Atacama Large Millimeter Array holds incredible promise for providing exquisite detail of the physical conditions in these extreme systems⁹. ■

SENSORY BIOLOGY

A whiff of genome

The relatedness of animals influences their behaviour, but how do they detect their kin? It seems that nasal sensory neurons differentiate between minute structural differences in urinary peptides that reflect genetic variation.

THOMAS BOEHM

Ask ten people how they would go about finding the right mate, and you will probably get ten different answers. But might they all be mistaken? Writing in *Nature Communications*, Sturm *et al.*¹ describe how the olfactory system is an amazingly sensitive and accurate discriminator of genetic relatedness, and suggest that this faculty may provide essential information that influences kin recognition, mate choice and other types of social interaction.

Many aspects of animal behaviour require information about the extent of genetic relatedness to other individuals. During the

Desika Narayanan is in the Department of Astronomy, University of Arizona, Tucson, Arizona 85721, USA, and the Department of Physics and Astronomy, Haverford College, Pennsylvania. Chris Carilli is at the National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, Socorro, New Mexico 87801, USA.

e-mails: dnarayanan@as.arizona.edu; ccarilli@nrao.edu

1. Riechers, D. A. *et al.* *Nature* **496**, 329–333 (2013).
2. Hughes, D. H. *et al.* *Nature* **394**, 241–247 (1998).
3. Barger, A. J. *et al.* *Nature* **394**, 248–251 (1998).
4. Chapman, S. C. *et al.* *Nature* **422**, 695–698 (2003).
5. Carilli, C. L. & Walter, F. *Annu. Rev. Astron. Astrophys.* (in the press); preprint at <http://arXiv.org/abs/1301.0371> (2013).
6. Vieira, J. D. *et al.* *Nature* **495**, 344–347 (2013).
7. Walter, F. *et al.* *Nature* **486**, 233–236 (2012).
8. Hayward, C. C. *et al.* *Mon. Not. R. Astron. Soc.* **428**, 2529–2547 (2013).
9. Blain, A. W. *Nature* **495**, 324–325 (2013).
10. Narayanan, D. *et al.* *Astrophys. J.* **176**, 331–354 (2008).

selection of the best possible mate, for example, some means of detecting family connections is essential to avoid inbreeding. Many sensory inputs contribute to solving this behavioural jigsaw puzzle, but the role of smell has received special attention. Volatile chemical compounds that attract potential mates over long distances are textbook examples of gender- and species-specific information carriers, but these pheromones convey no information on individuality. One possible source of such information is the thousands of DNA sequence differences present in animal genomes that allow even closely related individuals of a species to be easily distinguished². It seems that animals are capable of converting this genetic

code into olfactory information^{3,4}, but exactly how this is achieved is only just starting to become clear.

The genetic differences between individuals often manifest themselves in the altered arrangement of amino-acid residues in proteins. Could such alternative protein forms be converted into individuality signatures? At face value, proteins constitute unlikely olfactory stimuli because of their large size. But the proteins and small peptides (generated from protein breakdown) that are present in bodily fluids can become aerosolized and thereby gain direct access to the olfactory system.

In mice, the olfactory detection of two main types of genetically variable, or polymorphic, protein have been studied: the major urinary proteins (MUPs)³, which are excreted in copious amounts in the urine, and the proteins of the major histocompatibility complex (MHC)⁴, which are indispensable for effective immune responses. Indeed, controversy has arisen about which of these two classes of protein is responsible for individuality signals in mice^{5,6}.

Structural differences in MUPs are thought to be sensed directly by the olfactory system⁷, whereas it has been proposed⁴ that the structural diversity of MHC molecules is evaluated indirectly on the basis of the ligands that bind to them (Fig. 1). These ligands are short peptides that are breakdown products of intra- and extracellular proteins, and their structure reflects the binding properties of the MHC molecules, in a mode akin to a lock and key⁸. Hence, the 'MHC peptide hypothesis' suggests⁹ that olfactory sensing of the key (the peptide ligand) provides information about the lock (the MHC protein).

Peptides have been shown to activate olfactory sensory neurons at exceedingly low concentrations⁴, thereby satisfying one prediction of this hypothesis, but their presence in urine had not previously been demonstrated. To fill this void, Sturm *et al.* used mass spectrometry to search for urinary peptides that have characteristic sequences predicted to match the MHC molecules of mice with different genotypes. Although the authors found many such peptides, they made the surprising observation that the presence of these peptides was independent of MHC genotype. However, they also found that, in addition to MUPs, many other distinguishing genetic polymorphisms were reflected in the sequences of urinary peptides, suggesting that this complex 'peptidome' mirrors at least some part of genetic individuality.

But what about MHC peptide ligands that are predicted⁴ to also occur in the urine? Sturm and colleagues suggest that mass spectrometry is an unsuitable method to reproducibly detect low concentrations of specific MHC peptide ligands among the plethora of other urinary peptides present at million-fold higher concentrations. To get around this problem, the authors used 'empty' MHC molecules

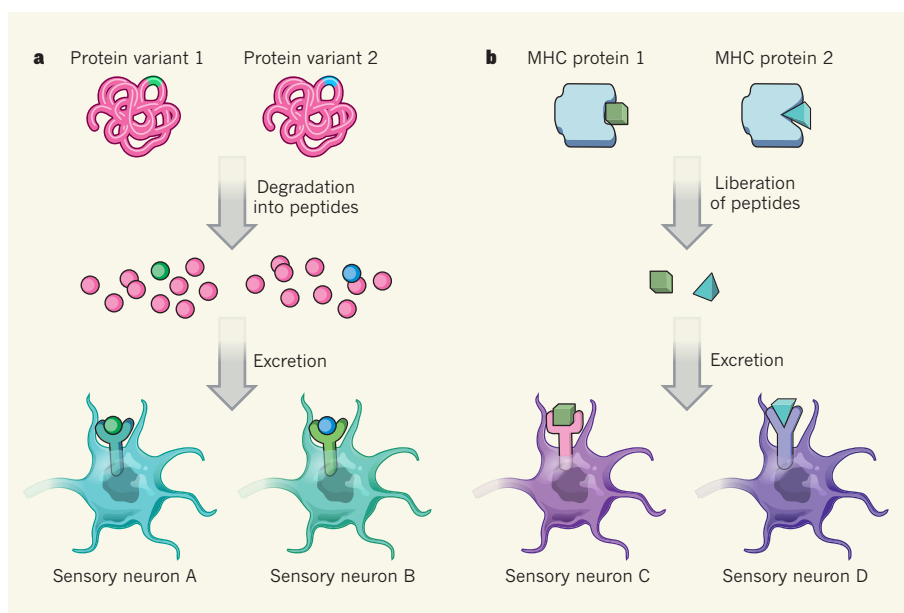


Figure 1 | Alternative sources of ligands for olfactory sensory neurons. **a**, Sturm *et al.*¹ show that degradation products of proteins are excreted into the urine and that these peptides serve as ligands for olfactory sensory neurons. Peptide variants that differ by a single amino-acid residue can activate different neurons, allowing the detection of genetic differences that lead to amino-acid alterations. **b**, The authors also show that peptides bound to MHC molecules activate olfactory neurons following liberation from the MHC and excretion. Because the structure of MHC molecules differs between individuals, the neuronal activation caused by these structurally diverse peptides will also reflect individuality.

to capture 'free' peptide ligands from mouse urine. By this means, they obtained unequivocal evidence that MHC ligands are indeed present in urine. They then studied the response of individual nasal sensory neurons to MHC-dependent and MHC-independent peptides, revealing that not only do individual neurons respond to these molecules at biologically relevant concentrations (ranging from 10^{-12} to 10^{-6} moles per litre), but they also faithfully discriminate between the two classes of peptide. Furthermore, the researchers observed distinct activation patterns in sensory neurons elicited by cocktails of peptides from mice of different genotypes, suggesting that these urinary peptides should thus be considered to constitute an 'individuality signature'.

These findings suggest that the two main competing hypotheses about how the olfactory system detects genetic individuality (through either MUPs or MHC ligands) may not be as mutually exclusive as was previously assumed. Instead, it seems that the sequence-dependent activation pattern of olfactory sensory neurons reflects the diverse origins of peptides contained in the urinary peptidome — ligands of polymorphic MHC molecules and breakdown products of other polymorphic proteins, perhaps including MUPs. The structural basis for the recognition of this diverse repertoire of peptide ligands remains unclear; one possibility is that, rather than expressing just one receptor, sensory neurons use hetero-oligomeric receptors for peptide detection.

From an evolutionary perspective, these ideas indicate a universal mechanism of

olfactory assessment of genetic individuality, even for animals that do not possess an MHC. But this work also has implications beyond animal behaviour. Olfactory recognition of MHC-bound peptides suggests that the exquisite sensitivity and precision of sensory evaluation might be used to detect the presence of infectious agents, because peptides derived from these agents are 'presented' to cells of the immune system by MHC molecules and might eventually appear in the urine. And from a more practical point of view, sensory neurons may now present themselves as a biological alternative to mass spectrometers, for use when the analysis of peptide mixtures depends on accuracy over a large range of concentrations. ■

Thomas Boehm is in the Department of Developmental Immunology, Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany.
e-mail: boehm@ie-freiburg.mpg.de

1. Sturm, T. *et al.* *Nature Commun.* <http://dx.doi.org/10.1038/ncomms2610> (2013).
2. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
3. Hurst, J. L. *et al.* *Nature* **414**, 631–634 (2001).
4. Leinders-Zufall, T. *et al.* *Science* **306**, 1033–1037 (2004).
5. Hurst, J. L. & Beynon, R. J. *BioEssays* **26**, 1288–1298 (2004).
6. Ruff, J. S., Nelson, A. C., Kubinak, J. L. & Potts, W. K. *Adv. Exp. Med. Biol.* **738**, 290–313 (2012).
7. Chamero, P. *et al.* *Nature* **450**, 899–902 (2007).
8. Rammensee, H.-G., Bachmann, J. & Stefanovic, S. *MHC Ligands and Peptide Motifs* (Landes Bioscience, 1997).
9. Boehm, T. *Nature Rev. Immunol.* **6**, 79–84 (2006).

Molecular motion watched

A laser pulse can switch certain crystals from an insulating phase to a highly conducting phase. The ultrafast molecular motions that drive the transition have been directly observed using electron diffraction. [SEE LETTER P.343](#)

BRADLEY SIWICK & ERIC COLLET

Richard Feynman once said, “everything that living things do can be understood in terms of the jiggling and wiggling of atoms”. It is not only biological processes, however, that are driven by the microscopic structural rearrangements of atoms and molecules — this idea cuts through many scientific disciplines. The underlying motions are usually hidden from view, because even the most advanced conventional methods have been unable to directly observe the dynamic evolution of atomic (or molecular) structure on its shortest timescales. But on page 343 of this issue, Gao *et al.*¹ report their use of ultrafast electron diffraction to provide a detailed look at one such hidden world. The authors have effectively ‘watched’ molecules rearrange in a crystal at the picosecond timescale on which chemical bonds are made and broken (1 picosecond (ps) is 10^{-12} seconds).

Specifically, Gao *et al.* have investigated a crystalline organic salt consisting of

electron-donor molecules (ethylenedioxy-tetrathiafulvalene, EDO-TTF) and hexafluorophosphate counterions (PF_6^-). At room temperature, the donor (D) molecules in the crystal are organized into linear chains that exhibit a high degree of ‘inversion’ symmetry. This high symmetry means that electrons are evenly shared by D molecules along the chain and that the material has high (metallic) conductivity (Fig. 1).

At temperatures below 280 kelvin, however, the electrons are trapped on some pairs of D molecules but not on others. The molecules without electrons become positively charged, giving rise to a periodic $\text{D}^+\text{D}^+\text{D}^0\text{D}^0$ molecular sequence along the chain (the numbers in the sequence indicate the charge on each donor). This lowers the symmetry of the system and prevents electrons from jumping easily from pair to pair, so that the crystal becomes an insulator. The salt also undergoes an ultrafast light-induced transition^{2,3} from the insulating phase to the metallic phase. Gao and colleagues set out to uncover the nature of the

light-induced molecular rearrangements that act as a switch for this transition.

Transitions like this one represent a new way in which matter can be manipulated by light⁴. In some materials, the subtle coupling between electronic and structural degrees of freedom following exposure to light seems to be strong enough to drive the system into an excited phase that has different electronic and structural order than the ground state, and which cannot be accessed through more conventional routes (for example, by using temperature, pressure, or electric or magnetic fields as control parameters). In order to control such ultrafast transitions for technological applications, the pathway along which structural and electronic changes occur must be observed in real time. How can this be accomplished?

Gao and co-workers’ answer for the $(\text{EDO-TTF})_2\text{PF}_6$ salt was to follow molecular motions through their effect on the electron-diffraction pattern of the crystal by combining tools and techniques from electron microscopy with those of ultrafast laser spectroscopy. The remarkable recent improvements in the spatial resolution of electron microscopes (now below 0.5 ångströms) have been paralleled by enhancements in the temporal resolution of both electron imaging and diffraction⁵. Gao and colleagues’ work raises the bar of achievement for electron-diffraction instrumentation, and was enabled by a series of breakthroughs that resulted from first understanding⁶ and then manipulating the evolution of ultrashort electron pulses using particle-accelerator technology^{7–9}.

The power of this technology is evident from the authors’ results. By collecting time-resolved electron-crystallography data of unprecedented quality from photoexcited $(\text{EDO-TTF})_2\text{PF}_6$, Gao *et al.* mapped the primary molecular motions involved in the earliest stages of the salt’s structural phase transition. These motions occur on timescales from a few hundred femtoseconds (fs; 1 femtosecond is 10^{-15} seconds) to several picoseconds, and were therefore previously uncharacterized. The authors compared the molecular arrangement of the salt’s unit cell at these early stages to the crystal structure of the insulating and metallic phases at thermal equilibrium. In this way, they determined that a transient intermediate structure unlike either equilibrium phase forms during the first few picoseconds after excitation.

These results add insight to what was known from previous spectroscopic studies^{2,3}. The researchers observed that, in the intermediate structure, molecules rearrange to form localized $\text{D}^+\text{D}^0\text{D}^+\text{D}^0$ sequences (Fig. 1). A second transformation then occurs to generate the long-range, high-symmetry stacking of the metallic structure. This second transformation was well characterized by the time course of symmetry-sensitive electron-diffraction peaks

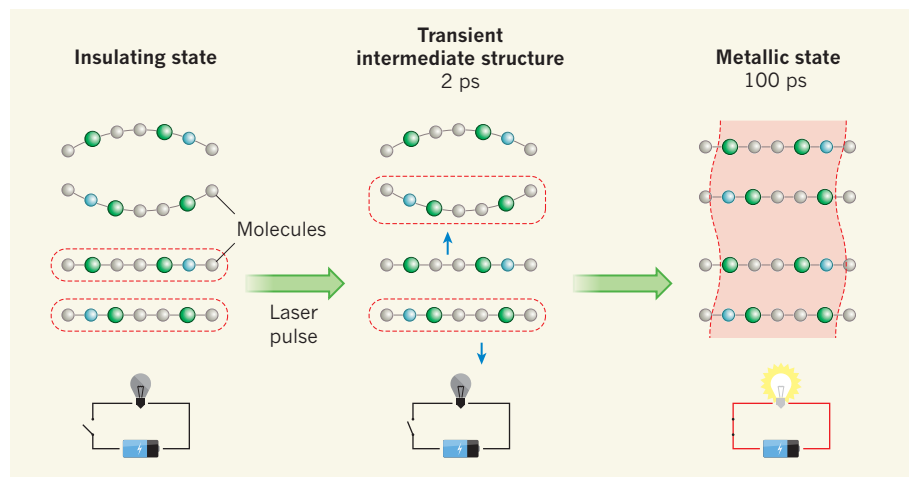


Figure 1 | Molecular motion of a phase switch. In the insulating state of the crystalline organic salt studied by Gao *et al.*¹, electron-donor molecules are either positively charged and straight, or neutral and bent. This reduces the symmetry of the system and traps positive charge (dashed red boxes) on pairs of adjacent molecules, preventing conduction — much like an open switch in a circuit (lower panel). In the salt’s metallic state, the symmetry of the system is higher because the molecules all carry equivalent charge, which delocalizes throughout the crystal (red shading), and allows conduction (closed switch and red wires in the circuit below). Gao *et al.* report that, 2 picoseconds after the insulating phase of the salt is excited by a laser pulse, a transient intermediate structure forms as a result of molecular motions (blue arrows). This intermediate contains localized structures in which electrons reside on alternate molecules. The intermediate state converts into the metallic state after 100 ps.

in Gao and colleagues' experiments, which indicates that the equilibrium metallic phase is reached within about 100 ps. Because conductivity correlates with crystal symmetry, the low-symmetry intermediate structure is more like an insulator than a metal.

Recent years have also seen tremendous success in the use of time-resolved X-ray diffraction to obtain atomic-level views of transient phenomena in inorganic, molecular and even biomolecular systems¹⁰, notably with the emergence of X-ray free-electron lasers¹¹. Gao and colleagues' work demonstrates that laboratory-scale, ultrafast electron-diffraction set-ups represent an attractive alternative to synchrotrons or X-ray free-electron lasers — massive installations that require electron accelerators kilometres in length. At present, however, the temporal resolution of the authors' apparatus is limited to about 400 fs, and so it could not catch the faster, coherent molecular motion (previously evidenced by optical spectroscopy²) that drives the system to the transient intermediate state itself. Nevertheless, their study represents an experimental tour de force.

Femtosecond electron diffraction opens up new possibilities for investigating weakly diffracting systems composed of light elements, such as organic materials and two-dimensional protein crystals, or for studying systems for which laser penetration is too small to perform

X-ray studies. More broadly, Gao and colleagues' work adds to the growing range of advanced structural studies, be they performed with electrons or X-rays, that are pushing the boundaries of experiments towards the limits of spatial and temporal resolution, opening the way towards greater understanding and control of molecular dynamics and material properties. ■

Bradley Siwick is in the Departments of Physics and Chemistry, and the Center for the Physics of Materials, McGill University, Montreal, Quebec, H3A 2K6 Canada.

Eric Collet is at the Institute of Physics, University of Rennes 1, 35042 Rennes, France. e-mails: bradley.siwick@mcgill.ca; eric.collet@univ-rennes1.fr

1. Gao, M. *et al.* *Nature* **496**, 343–346 (2013).
2. Chollet, M. *et al.* *Science* **307**, 86–89 (2005).
3. Onda, K. *et al.* *Phys. Rev. Lett.* **101**, 067403 (2008).
4. Nasu, K. (ed.) *Photoinduced Phase Transitions* (World Scientific, 2004).
5. Sciaini, G. *et al.* *Rep. Prog. Phys.* **74**, 096101 (2011).
6. Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Dwayne Miller, R. J. *J. Appl. Phys.* **92**, 1643–1648 (2002).
7. van Oudheusden, T. *et al.* *J. Appl. Phys.* **102**, 093501 (2007).
8. Gao, M. *et al.* *Opt. Express* **20**, 12048–12058 (2012).
9. Chatelain, R. P., Morrison, V. R., Godbout, C. & Siwick, B. J. *J. Appl. Phys. Lett.* **101**, 081901 (2012).
10. Collet, E. *Acta Crystallogr. A* **66**, 133–134 (2010).
11. Harmand, M. *et al.* *Nature Photonics* **7**, 215–218 (2013).

BIOGEOCHEMISTRY

Nitrogen deposition and forest carbon

Human activity, such as agricultural fertilizer use, has increased the amount of nitrogen deposited onto forests from the atmosphere. The photosynthetic response to this in evergreen needleleaf forests has been quantified globally.

BEVERLY LAW

Nitrogen-containing compounds deposited from the atmosphere can affect the amount of carbon that is absorbed into ecosystems by photosynthesis. Writing in *Global Biogeochemical Cycles*, Fleischer *et al.*¹ reveal that photosynthesis by boreal and temperate evergreen forests rises with increased atmospheric nitrogen deposition, but levels out when a threshold value of 8 kilograms of nitrogen per hectare per year is reached. These findings highlight the need to clarify the connections between carbon and nitrogen in the environment, and to disentangle the effects of climate from those of nitrogen deposition on forests.

The terrestrial biosphere is thought to take

up about 30% of human-produced carbon dioxide from the atmosphere annually, lessening the greenhouse effect of fossil-fuel emissions². However, estimates of the size of this terrestrial sink are uncertain because of major gaps in our knowledge of the magnitude of the effects of the factors involved. One of these factors is the variation in human-caused additions of nitrogen to the atmosphere over time and space. Climate change and atmospheric fertilization of plants by human sources of nitrogen and CO₂ probably affect plant growth rates all over the world, but our understanding of these effects is likely to remain incomplete for the foreseeable future³.

Most plants cannot use nitrogen gas in the atmosphere for growth. They require it to be converted to usable forms, such as

ammonia — a process that forms part of the nitrogen cycle. However, the natural nitrogen cycle has been heavily influenced by human activities, which produce highly unstable forms of nitrogen known collectively as reactive nitrogen. The deposition of reactive nitrogen from the atmosphere to forests occurs mainly as a result of agricultural fertilizer use and fossil-fuel combustion, and has increased from 15 million tonnes of human-produced reactive nitrogen per year in the 1860s to 187 million tonnes by 2005 (ref. 4). Nitrogen deposition is expected to continue to increase in many regions, and has been predicted⁵ to almost double globally by 2050.

The addition of reactive nitrogen to the atmosphere by humans affects climate, and the composition and function of terrestrial and aquatic ecosystems⁴. High levels of nitrogen deposition have many damaging effects on ecosystems, but small additions can be beneficial for otherwise nitrogen-limited ecosystems, because they increase the uptake of CO₂ from the atmosphere by photosynthesis. The effect of nitrogen deposition on carbon sequestration by soil is less clear. Most studies suggest that the net effect is between 35 and 65 kilograms of carbon sequestered per kilogram of nitrogen^{6,7}. Much of the variation can be attributed to different patterns of growth and to the availability of other resources for growth.

Although low levels of nitrogen deposition might mitigate the effects of increased atmospheric CO₂ to some degree, 53–76% of this coincidental benefit is itself estimated to be offset globally⁷. This is because nitrogen deposition can stimulate net emissions of other greenhouse gases (methane and nitrous oxide) that are products of microbial activity in the soils of many ecosystems. Thus, the relative contribution of nitrogen deposition to the strength of the terrestrial carbon sink remains uncertain.

To address this problem, near-continuous atmospheric observations have been made from towers above vegetation canopies to provide estimates of CO₂ uptake (photosynthesis) and release (respiration) from and to the atmosphere, respectively. Many of these sites have been running for more than a decade, and the data are summarized by the FLUXNET Project⁸. In combination with biological and environmental data, such information has been used by researchers to examine the effect of factors such as climate variability on carbon processes at the whole-ecosystem scale at many sites around the world.

Fleischer *et al.* used FLUXNET data from 80 sites that had sufficient information about nitrogen and carbon fluxes for their analysis. They found that, for evergreen needleleaf forests in temperate and boreal zones, maximal photosynthesis under optimal environmental conditions increased with continuous nitrogen-deposition rates up to a threshold of about 8 kg of nitrogen deposition per hectare

per year. Above this value, no further increase in photosynthesis was observed.

Forests above the threshold are therefore at an intermediate stage of nitrogen saturation — a stage at which nitrogen availability exceeds microbial and plant demands, and can result in some nitrogen leaching from the ecosystem. Prolonged availability of excessive nitrogen can lead to more leaching, decreased growth and nutrient imbalances. By contrast, the evergreen needleleaf forests that responded most strongly to nitrogen deposition are in a nitrogen-limited range, within which photosynthetic capacity increases with deposition. The observed threshold is a small fraction of the nitrogen that farmers would use annually in fertilizers.

The authors also found that boreal evergreen needleleaf forests had a slightly lower photosynthetic response overall to nitrogen deposition. For both boreal and temperate evergreen forests, this translates to roughly 25 kg of carbon sequestered per kilogram of nitrogen, less than the estimated global average^{6,7,9} of 35–65.

The results are confounded by the effects of climate on photosynthesis: the nitrogen-deposition effect may be larger or smaller than Fleischer and colleagues' findings because part of the observed response is probably a result of climate. However, there is no evidence to suggest that the contribution of nitrogen deposition is zero. Although the authors attempted to determine the thresholds for other forest types, they were limited by the availability of biological data on both carbon and nitrogen processes. More comprehensive measurements of nitrogen stocks and cycling at the global network of carbon monitoring sites are required

to separate the effects of climate and nitrogen deposition on forests.

The net emissions of all greenhouse gases — including CO₂, methane and nitrous oxide — should be considered when examining the net effects of nitrogen deposition and climate on ecosystems. A better understanding of how the connections between carbon and nitrogen in the environment could change in the future is also required. Nevertheless, Fleischer and co-workers' study lays the groundwork needed to refine estimates of the effects of climate and nitrogen deposition on the terrestrial biosphere's ability to remove carbon from the atmosphere as CO₂ emissions increase. Such refinements will be necessary to improve predictions of the effects of these emissions on ecosystems at local, regional and continental scales. ■

Beverly Law is in the Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon 97331, USA. e-mail: bev.law@oregonstate.edu

1. Fleischer, K. *et al.* *Glob. Biogeochem. Cycl.* <http://dx.doi.org/10.1002/gbc.20026> (2013).
2. Canadell, J. G. *et al.* *Proc. Natl Acad. Sci. USA* **104**, 18866–18870 (2007).
3. National Research Council. *Verifying Greenhouse Gas Emissions* (National Academies Press, 2010).
4. Galloway, J. N. *et al.* *Science* **320**, 889–892 (2008).
5. Galloway, J. N. *et al.* *Biogeochemistry* **70**, 153–226 (2004).
6. Butterbach-Bahl, K. *et al.* in *The European Nitrogen Assessment* (eds Sutton, M. *et al.*) 99–125 (Cambridge Univ. Press, 2011).
7. Liu, L. & Greaver, T. L. *Ecol. Lett.* **12**, 1103–1117 (2009).
8. Baldocchi, D. D. *et al.* *Bull. Am. Meteorol. Soc.* **82**, 2415–2434 (2001).
9. Erisman, J. W. *et al.* *Curr. Opin. Environ. Sustainability* **3**, 281–290 (2011).

COMPLEX SYSTEMS

Spatial signatures of resilience

Predicting when the dynamics of a complex system will undergo a sudden transition is difficult. New experiments show that the spatial distribution of organisms can indicate when such tipping points are near. SEE LETTER P.355

STEPHEN R. CARPENTER

The divergence of dynamics towards sharply different states occurs in complex systems in fields ranging from physics and physiology to ecology and social sciences. The thresholds for these critical transitions are often unknown until surprising shifts occur. Establishing measures of the distance to a threshold — the resilience of a system — could allow researchers to compare

the stability of different systems or even anticipate an impending transition. In a paper on page 355 of this issue, Dai *et al.*¹ present a novel resilience index and use ingenious laboratory experiments to support the theory underlying it*. Their approach is based on the spatial distribution of organisms and thereby adds, quite literally, a new dimension to attempts to predict transitions.

*This article and the paper under discussion were published online on 10 April 2013.

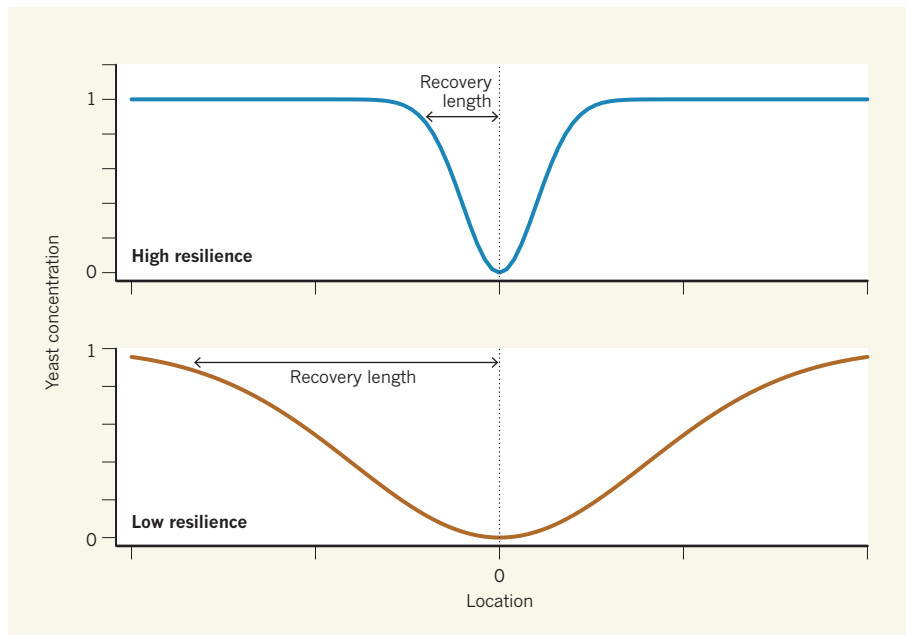


Figure 1 | Paths to recovery. Dai and colleagues¹ studied a series of linked cultures of yeast populations subject to catastrophic collapse when the dilution factor is raised above a critical threshold. They then introduced a 'bad patch' (position 0) in the form of a culture with a high dilution factor. Recovery length is the distance that a population must be from a bad patch for it to recover 90% of its original high concentration (a population size of 1). When the entire series of cultures was highly resilient (far from the threshold) the recovery length was short (top panel), whereas when the series of cultures had low resilience (close to the threshold) the recovery length was long (bottom panel).

Examples of the spatial distribution of organisms include the pattern of trees in a forest, of aquatic insects on the bottom of a stream, or of mussels along a rocky shoreline. In such ecosystems, a patch of unusually good habitat will be surrounded by a halo of patches with high population density, owing to the diffusion of organisms from the good patch to neighbouring sites. Likewise, a patch of poor habitat creates a halo of low population density. As the distance away from the patch of unusual habitat increases, the population density will gradually return to the average for the landscape as a whole. The distance required for the population density to return to the average value is referred to as the recovery length. Dai and colleagues show that the recovery length provides an indication of the resilience of the landscape pattern (Fig. 1).

The proposal to use recovery length as a resilience index is in contrast to most research on resilience indicators, which has used time series without spatial information², such as populations of fish, density of grasses, or algae concentrations in a lake. However, high-resolution time series are rare in ecology and may not be available for crucial ecosystems when managers need to make decisions. Spatial snapshots of resilience offer unique advantages³: they do not require intensive time series, and they may often be reconstructed from archives of aerial photos or satellite images.

To demonstrate the usefulness of recovery length, Dai *et al.* designed a system of yeast cultures that were physically linked and that exhibited alternate stable states of high or low

cell concentration, to represent patches in an ecosystem. In their experiments, the proportion of the liquid culture medium replaced each day (the dilution factor) had a critical threshold. At a low dilution factor, the yeast concentration of a patch was at the high steady state. As the dilution factor was gradually increased, the yeast concentration of the patch fell smoothly until the threshold dilution rate was reached, and then fell sharply to the low steady state. Thus, the gradual increase in dilution rate from a low value erodes the resilience of the system until the threshold is reached and the yeast population collapses.

The experimenters introduced a 'bad patch' (with a high dilution factor and therefore a low yeast concentration) into a series of linked cultures with identical dilution factors and yeast concentrations near the high steady state. As distance away from the bad patch increased, the yeast concentration increased gradually towards the high steady state (Fig. 1). The distance required for the yeast population to recover to the high steady state is the recovery length. As the dilution factor, and therefore the resilience, of the entire series of cultures declined, the recovery length increased.

Previous studies have shown that populations become more variable and recover more slowly from disturbance as resilience declines^{3,4}. But Dai and colleagues demonstrate that these indicators of resilience are dampened by movements of organisms between patches. Furthermore, their results indicate that the influence of a patch on its neighbours

becomes more widespread as resilience declines (Fig. 1). So, as resilience decreases, the effects of a bad patch spread farther from the epicentre. Similar effects are known from spatial models of populations subjected to harvesting⁵. Increases in recovery length may also be related to the decreasing abundance of large patches, as occurs in real landscapes when overgrazing or climate change transforms grasslands into deserts⁶.

This study and others^{7,8} demonstrate how laboratory experiments can connect model studies of resilience to real-world field applications. Simulation models are rapidly expanding our knowledge of the situations in which resilience indicators can or cannot be used, and of the statistical challenges of measuring resilience in specific situations. However, few field experiments use manipulated and reference ecosystems to evaluate resilience indicators⁹. We need intensive high-quality data collected in multiple ecosystems for many years before we will be able to measure resilience at the scales needed for ecosystem management.

Dai and colleagues' results suggest that assessments of recovery length may be a powerful approach for determining resilience in large-scale field studies. According to their findings, in resilient ecosystems, a patch of unusually good or bad quality created by experimenters should have limited spatial influence. But in an ecosystem close to collapse, a patch of unusual quality should have widespread effects. Instead of collecting detailed time series of population densities through laborious field work, researchers can assess resilience by measuring recovery length in maps derived from remotely sensed data at just a few time points.

Thresholds for the collapse of harvested populations, the degradation of rangelands, toxic algal blooms in water supplies, species invasions and population extinctions will continue to challenge ecologists and natural-resource managers. We will never eliminate surprises. But recovery length may provide a tool that allows us to steer away from some thresholds before they are crossed. ■

Stephen R. Carpenter is at the Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.
e-mail: srcarpen@wisc.edu

1. Dai, L., Korolev, K. S. & Gore, J. *Nature* **496**, 355–358 (2013).
2. Dakos, V. *et al.* *PLoS ONE* **7**, e41010 (2012).
3. Scheffer, M. *et al.* *Science* **338**, 344–348 (2012).
4. Scheffer, M. *et al.* *Nature* **461**, 53–59 (2009).
5. Carpenter, S. R. & Brock, W. A. *Ecosphere* **1**, art10 (2010).
6. Kéfi, S. *et al.* *Nature* **449**, 213–217 (2007).
7. Drake, J. M. & Griffen, B. D. *Nature* **467**, 456–459 (2010).
8. Veraart, A. J. *et al.* *Nature* **481**, 357–359 (2012).
9. Carpenter, S. R. *et al.* *Science* **332**, 1079–1082 (2011).

The African coelacanth genome provides insights into tetrapod evolution

Chris T. Amemiya^{1,2*}, Jessica Alföldi^{3*}, Alison P. Lee⁴, Shaohua Fan⁵, Hervé Philippe⁶, Iain MacCallum³, Ingo Braasch⁷, Tereza Manousaki^{5,8}, Igor Schneider⁹, Nicolas Rohner¹⁰, Chris Organ¹¹, Domitille Chalopin¹², Jeramiah J. Smith¹³, Mark Robinson¹, Rosemary A. Dorrington¹⁴, Marco Gerdo¹⁵, Bronwen Aken¹⁶, Maria Assunta Biscotti¹⁷, Marco Barucca¹⁷, Denis Baurain¹⁸, Aaron M. Berlin³, Gregory L. Blatch^{14,19}, Francesco Buonocore²⁰, Thorsten Burmester²¹, Michael S. Campbell²², Adriana Canapa¹⁷, John P. Cannon²³, Alan Christoffels²⁴, Gianluca De Moro¹⁵, Adrienne L. Edkins¹⁴, Lin Fan³, Anna Maria Fausto²⁰, Nathalie Feiner^{5,25}, Mariko Forconi¹⁷, Junaid Gamielien²⁴, Sante Gnerre³, Andreas Gnirke³, Jared V. Goldstone²⁶, Wilfried Haerty²⁷, Mark E. Hahn²⁶, Uljana Hesse²⁴, Steve Hoffmann²⁸, Jeremy Johnson³, Sibel I. Karchner²⁶, Shigehiro Kuraku^{5†}, Marcia Lara³, Joshua Z. Levin³, Gary W. Litman²³, Evan Mauceli^{3†}, Tsutomu Miyake²⁹, M. Gail Mueller³⁰, David R. Nelson³¹, Anne Nitsche³², Ettore Olmo¹⁷, Tatsuya Ota³³, Alberto Pallavicini¹⁵, Sumir Panji^{24†}, Barbara Picone²⁴, Chris P. Ponting²⁷, Sonja J. Prohaska³⁴, Dariusz Przybylski³, Nil Ratan Saha¹, Vydianathan Ravi⁴, Filipe J. Ribeiro^{3†}, Tatjana Sauka-Spengler³⁵, Giuseppe Scapigliati²⁰, Stephen M. J. Searle¹⁶, Ted Sharpe³, Oleg Simakov^{5,36}, Peter F. Stadler³², John J. Stegeman²⁶, Kenta Sumiyama³⁷, Diana Tabbaa³, Hakim Tafer³², Jason Turner-Maier³, Peter van Heusden²⁴, Simon White¹⁶, Louise Williams³, Mark Yandell²², Henner Brinkmann⁶, Jean-Nicolas Volff¹², Clifford J. Tabin¹⁰, Neil Shubin³⁸, Manfred Schartl³⁹, David B. Jaffe³, John H. Postlethwait⁷, Byrappa Venkatesh⁴, Federica Di Palma³, Eric S. Lander³, Axel Meyer^{5,8,25} & Kerstin Lindblad-Toh^{3,40}

The discovery of a living coelacanth specimen in 1938 was remarkable, as this lineage of lobe-finned fish was thought to have become extinct 70 million years ago. The modern coelacanth looks remarkably similar to many of its ancient relatives, and its evolutionary proximity to our own fish ancestors provides a glimpse of the fish that first walked on land. Here we report the genome sequence of the African coelacanth, *Latimeria chalumnae*. Through a phylogenomic analysis, we conclude that the lungfish, and not the coelacanth, is the closest living relative of tetrapods. Coelacanth protein-coding genes are significantly more slowly evolving than those of tetrapods, unlike other genomic features. Analyses of changes in genes and regulatory elements during the vertebrate adaptation to land highlight genes involved in immunity, nitrogen excretion and the development of fins, tail, ear, eye, brain and olfaction. Functional assays of enhancers involved in the fin-to-limb transition and in the emergence of extra-embryonic tissues show the importance of the coelacanth genome as a blueprint for understanding tetrapod evolution.

In 1938 Marjorie Courtenay-Latimer, the curator of a small natural history museum in East London, South Africa, discovered a large, unusual-looking fish among the many specimens delivered to her by a local fish trawler. *Latimeria chalumnae*, named after its discoverer¹, was over 1 m long, bluish in colour and had conspicuously fleshy fins that resembled the limbs of terrestrial vertebrates. This discovery is

considered to be one of the most notable zoological finds of the twentieth century. *Latimeria* is the only living member of an ancient group of lobe-finned fishes that was known previously only from fossils and believed to have been extinct since the Late Cretaceous period, approximately 70 million years ago (Myr ago)¹. It was almost 15 years before a second specimen of this elusive species was discovered in the

¹Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington 98101, USA. ²Department of Biology, University of Washington, Seattle, Washington 98105, USA. ³Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁴Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673, Singapore. ⁵Department of Biology, University of Konstanz, Konstanz 78464, Germany. ⁶Département de Biochimie, Université de Montréal, Centre Robert Cederghen, Montréal H3T 1J4, Canada. ⁷Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, USA. ⁸Konstanz Research School of Chemical Biology, University of Konstanz, Konstanz 78464, Germany. ⁹Instituto de Ciencias Biológicas, Universidade Federal do Para, Belem 66075-110, Brazil. ¹⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹Department of Anthropology, University of Utah, Salt Lake City, Utah 84112, USA. ¹²Institut de Genomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, Lyon 69007, France. ¹³Department of Biology, University of Kentucky, Lexington, Kentucky 40506, USA. ¹⁴Biomedical Biotechnology Research Unit (BioBRU), Department of Biochemistry, Microbiology & Biotechnology, Rhodes University, Grahamstown 6139, South Africa. ¹⁵Department of Life Sciences, University of Trieste, Trieste 34128, Italy. ¹⁶Department of Informatics, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ¹⁷Department of Life and Environmental Sciences, Polytechnic University of Marche, Ancona 60131, Italy. ¹⁸Department of Life Sciences, University of Liege, Liege 4000, Belgium. ¹⁹College of Health and Biomedicine, Victoria University, Melbourne VIC 8001, Australia. ²⁰Department for Innovation in Biological, Agro-food and Forest Systems, University of Tuscia, Viterbo 01100, Italy. ²¹Department of Biology, University of Hamburg, Hamburg 20146, Germany. ²²Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA. ²³Department of Pediatrics, University of South Florida Morsani College of Medicine, Children's Research Institute, St. Petersburg, Florida 33701, USA. ²⁴South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, South Africa. ²⁵International Max-Planck Research School for Organismal Biology, University of Konstanz, Konstanz 78464, Germany. ²⁶Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. ²⁷MRC Functional Genomics Unit, Oxford University, Oxford OX1 3PT, UK. ²⁸Transcriptome Bioinformatics Group, LIFE Research Center for Civilization Diseases, Universität Leipzig, Leipzig 04109, Germany. ²⁹Graduate School of Science and Technology, Keio University, Yokohama 223-8522, Japan. ³⁰Department of Molecular Genetics, All Children's Hospital, St. Petersburg, Florida 33701, USA. ³¹Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA. ³²Bioinformatics Group, Department of Computer Science, Universität Leipzig, Leipzig 04109, Germany. ³³Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies, Hayama 240-0193, Japan. ³⁴Computational EvoDevo Group, Department of Computer Science, Universität Leipzig, Leipzig 04109, Germany. ³⁵Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX1 2JD, UK. ³⁶European Molecular Biology Laboratory, Heidelberg 69117, Germany. ³⁷Division of Population Genetics, National Institute of Genetics, Mishima 411-8540, Japan. ³⁸University of Chicago, Chicago, Illinois 60637, USA. ³⁹Department Physiological Chemistry, Biocenter, University of Würzburg, Würzburg 97070, Germany. ⁴⁰Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden. †Present addresses: Genome Resource and Analysis Unit, Center for Developmental Biology, RIKEN, Kobe, Japan (S.K.); Boston Children's Hospital, Boston, Massachusetts, USA (E.M.); Computational Biology Unit, Institute of Infectious Disease and Molecular Medicine, University of Cape Town Health Sciences Campus, Anzio Road, Observatory 7925, South Africa (S.P.); New York Genome Center, New York, New York, USA (F.J.R.).

*These authors contributed equally to this work.

Comoros Islands in the Indian Ocean, and only 309 individuals have been recorded in the past 75 years (R. Nulens, personal communication)². The discovery in 1997 of a second coelacanth species in Indonesia, *Latimeria menadoensis*, was equally surprising, as it had been assumed that living coelacanths were confined to small populations off the East African coast^{3,4}. Fascination with these fish is partly due to their prehistoric appearance—remarkably, their morphology is similar to that of fossils that date back at least 300 Myr, leading to the supposition that, among vertebrates, this lineage is markedly slow to evolve^{1,5}. *Latimeria* has also been of particular interest to evolutionary biologists, owing to its hotly debated relationship to our last fish ancestor, the fish that first crawled onto land⁶. In the past 15 years, targeted sequencing efforts have produced the sequences of the coelacanth mitochondrial genomes⁷, HOX clusters⁸ and a few gene families^{9,10}. Nevertheless, coelacanth research has felt the lack of large-scale sequencing data. Here we describe the sequencing and comparative analysis of the genome of *L. chalumnae*, the African coelacanth.

Genome assembly and annotation

The African coelacanth genome was sequenced and assembled using DNA from a Comoros Islands *Latimeria chalumnae* specimen (Supplementary Fig. 1). It was sequenced by Illumina sequencing technology and assembled using the short read genome assembler ALLPATHS-LG¹¹. The *L. chalumnae* genome has been reported previously to have a karyotype of 48 chromosomes¹². The draft assembly is 2.86 gigabases (Gb) in size and is composed of 2.18 Gb of sequence plus gaps between contigs. The coelacanth genome assembly has a contig N50 size (the contig size above which 50% of the total length of the sequence assembly can be found) of 12.7 kilobases (kb) and a scaffold N50 size of 924 kb, and quality metrics comparable to other Illumina genomes (Supplementary Note 1, and Supplementary Tables 1 and 2).

The genome assembly was annotated separately by both the Ensembl gene annotation pipeline (Ensembl release 66, February 2012) and by MAKER¹³. The Ensembl gene annotation pipeline created gene models using protein alignments from the Universal Protein Resource (UniProt) database, limited coelacanth complementary DNA data, RNA-seq data generated from *L. chalumnae* muscle (18 Gb of paired-end reads were assembled using Trinity software¹⁴, Supplementary Fig. 2) as well as orthology with other vertebrates. This pipeline produced 19,033 protein-coding genes containing 21,817 transcripts. The MAKER pipeline used the *L. chalumnae* Ensembl gene set, Uniprot protein alignments, and *L. chalumnae* (muscle) and *L. menadoensis* (liver and testis)¹⁵ RNA-seq data to create gene models, and this produced 29,237 protein-coding gene annotations. In addition, 2,894 short non-coding RNAs, 1,214 long non-coding RNAs, and more than 24,000 conserved RNA secondary structures were identified (Supplementary Note 2, Supplementary Tables 3 and 4, Supplementary Data 1–3 and Supplementary Fig. 3). It was inferred that 336 genes underwent specific duplications in the coelacanth lineage (Supplementary Note 3, Supplementary Tables 5 and 6, and Supplementary Data 4).

The closest living fish relative of tetrapods

The question of which living fish is the closest relative to ‘the fish that first crawled on to land’ has long captured our imagination: among scientists the odds have been placed on either the lungfish or the coelacanth¹⁶. Analyses of small to moderate amounts of sequence data for this important phylogenetic question (ranging from 1 to 43 genes) has tended to favour the lungfishes as the extant sister group to the land vertebrates¹⁷. However, the alternative hypothesis that the lungfish and the coelacanth are equally closely related to the tetrapods could not be rejected with previous data sets¹⁸.

To seek a comprehensive answer we generated RNA-seq data from three samples (brain, gonad and kidney, and gut and liver) from the West African lungfish, *Protopterus annectens*, and compared it to gene sets from 21 strategically chosen jawed vertebrate species. To perform a reliable analysis we selected 251 genes in which a 1:1 orthology ratio

was clear and used CAT-GTR, a complex site-heterogeneous model of sequence evolution that is known to reduce tree-reconstruction artefacts¹⁹ (see Supplementary Methods). The resulting phylogeny, based on 100,583 concatenated amino acid positions (Fig. 1, posterior probability = 1.0 for the lungfish–tetrapod node) is maximally supported except for the relative positions of the armadillo and the elephant. It corroborates known vertebrate phylogenetic relationships and strongly supports the conclusion that tetrapods are more closely related to lungfish than to the coelacanth (Supplementary Note 4 and Supplementary Fig. 4).

The slowly evolving coelacanth

The morphological resemblance of the modern coelacanth to its fossil ancestors has resulted in it being nicknamed ‘the living fossil’¹. This invites the question of whether the genome of the coelacanth is as slowly evolving as its outward appearance suggests. Earlier work showed that a few gene families, such as Hox and protocadherins, have comparatively slower protein-coding evolution in coelacanth than in other vertebrate lineages^{8,10}. To address the question, we compared several features of the coelacanth genome to those of other vertebrate genomes.

Protein-coding gene evolution was examined using the phylogenomics data set described above (251 concatenated proteins) (Fig. 1). Pair-wise distances between taxa were calculated from the branch lengths of the tree using the two-cluster test proposed previously²⁰ to test for equality of average substitution rates. Then, for each of the following species and species clusters (coelacanth, lungfish, chicken and mammals), we ascertained their respective mean distance to an outgroup consisting of three cartilaginous fishes (elephant shark, little skate and spotted catshark). Finally, we tested whether there was any significant difference in the distance to the outgroup of cartilaginous fish for every pair of species and species clusters, using a

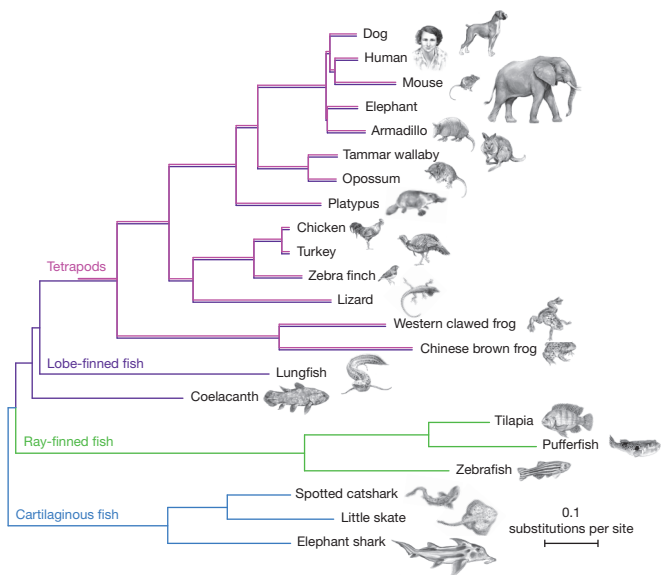


Figure 1 | A phylogenetic tree of a broad selection of jawed vertebrates shows that lungfish, not coelacanth, is the closest relative of tetrapods.

Multiple sequence alignments of 251 genes with a 1:1 ratio of orthologues in 22 vertebrates and with a full sequence coverage for both lungfish and coelacanth were used to generate a concatenated matrix of 100,583 unambiguously aligned amino acid positions. The Bayesian tree was inferred using PhyloBayes under the CAT + GTR + Γ_4 model with confidence estimates derived from 100 gene jack-knife replicates (support is 100% for all clades but armadillo + elephant with 45%)⁴⁸. The tree was rooted on cartilaginous fish, and shows that the lungfish is more closely related to tetrapods than the coelacanth, and that the protein sequence of coelacanth is evolving slowly. Pink lines (tetrapods) are slightly offset from purple lines (lobe-finned fish), to indicate that these species are both tetrapods and lobe-finned fish.

Z statistic. When these distances to the outgroup of cartilaginous fish were compared, we found that the coelacanth proteins that were tested were significantly more slowly evolving (0.890 substitutions per site) than the lungfish (1.05 substitutions per site), chicken (1.09 substitutions per site) and mammalian (1.21 substitutions per site) orthologues ($P < 10^{-6}$ in all cases) (Supplementary Data 5). In addition, as can be seen in Fig. 1, the substitution rate in coelacanth is approximately half that in tetrapods since the two lineages diverged. A Tajima's relative rate test²¹ confirmed the coelacanth's significantly slower rate of protein evolution ($P < 10^{-20}$) (Supplementary Data 6).

We next examined the abundance of transposable elements in the coelacanth genome. Theoretically, transposable elements may make their greatest contribution to the evolution of a species by generating templates for exaptation to form novel regulatory elements and exons, and by acting as substrates for genomic rearrangement²². We found that the coelacanth genome contains a wide variety of transposable-element superfamilies and has a relatively high transposable-element content (25%); this number is probably an underestimate as this is a draft assembly (Supplementary Note 5 and Supplementary Tables 7–10). Analysis of RNA-seq data and of the divergence of individual transposable-element copies from consensus sequences show that 14 coelacanth transposable-element superfamilies are currently active (Supplementary Note 6, Supplementary Table 10 and Supplementary Fig. 5). We conclude that the current coelacanth genome shows both an abundance and activity of transposable elements similar to many other genomes. This contrasts with the slow protein evolution observed.

Analyses of chromosomal breakpoints in the coelacanth genome and tetrapod genomes reveal extensive conservation of synteny and indicate that large-scale rearrangements have occurred at a generally low rate in the coelacanth lineage. Analyses of these rearrangement classes detected several fission events published previously²³ that are known to have occurred in tetrapod lineages, and at least 31 inter-chromosomal rearrangements that occurred in the coelacanth lineage or the early tetrapod lineage (0.063 fusions per 1 Myr), compared to 20 events (0.054 fusions per 1 Myr) in the salamander lineage and 21 events (0.057 fusions per 1 Myr) in the *Xenopus* lineage²³ (Supplementary Note 7 and Supplementary Fig. 6). Overall, these analyses indicate that karyotypic evolution in the coelacanth lineage has occurred at a relatively slow rate, similar to that of non-mammalian tetrapods²⁴.

In a separate analysis we also examined the evolutionary divergence between the two species of coelacanth, *L. chalumnae* and *L. menadoensis*, found in African and Indonesian waters, respectively. Previous analysis of mitochondrial DNA showed a sequence identity of 96%, but estimated divergence times range widely from 6 to 40 Myr^{25,26}. When we compared the liver and testis transcriptomes of *L. menadoensis*²⁷ to the *L. chalumnae* genome, we found an identity of 99.73% (Supplementary Note 8 and Supplementary Fig. 7), whereas alignments between 20 sequenced *L. menadoensis* bacterial artificial chromosomes (BACs) and the *L. chalumnae* genome showed an identity of 98.7% (Supplementary Table 11 and Supplementary Fig. 8). Both the genic and genomic divergence rates are similar to those seen between the human and chimpanzee genomes (99.5% and 98.8%, respectively; divergence time of 6 to 8 Myr ago)²⁸, whereas the rates of molecular evolution in *Latimeria* are probably affected by several factors, including the slower substitution rate seen in coelacanth. This suggests a slightly longer divergence time for the two coelacanth species.

The adaptation of vertebrates to land

As the species with a sequenced genome closest to our most recent aquatic ancestor, the coelacanth provides a unique opportunity to identify genomic changes that were associated with the successful adaptation of vertebrates to the land environment.

Over the 400 Myr that vertebrates have lived on land, some genes that are unnecessary for existence in their new environment have been eliminated. To understand this aspect of the water-to-land transition,

we surveyed the *Latimeria* genome annotations to identify genes that were present in the last common ancestor of all bony fish (including the coelacanth) but that are missing from tetrapod genomes. More than 50 such genes, including components of fibroblast growth factor (FGF) signalling, TGF- β and bone morphogenic protein (BMP) signalling, and WNT signalling pathways, as well as many transcription factor genes, were inferred to be lost based on the coelacanth data (Supplementary Data 7 and Supplementary Fig. 9). Previous studies of genes that were lost in this transition could only compare teleost fish to tetrapods, meaning that differences in gene content could have been due to loss in the tetrapod or in the lobe-finned fish lineages. We were able to confirm that four genes that were shown previously to be absent in tetrapods (*And1* and *And2* (ref. 29), *Fgf24* (ref. 30) and *Asip2* (ref. 31)), were indeed present and intact in *Latimeria*, supporting the idea that they were lost in the tetrapod lineage.

We functionally annotated more than 50 genes lost in tetrapods using zebrafish data (gene expression, knock-downs and knockouts). Many genes were classified in important developmental categories (Supplementary Data 7): fin development (13 genes); otolith and ear development (8 genes); kidney development (7 genes); trunk, somite and tail development (11 genes); eye (13 genes); and brain development (23 genes). This implies that critical characters in the morphological transition from water to land (for example, fin-to-limb transition and remodelling of the ear) are reflected in the loss of specific genes along the phylogenetic branch leading to tetrapods. However, homeobox genes, which are responsible for the development of an organism's basic body plan, show only slight differences between *Latimeria*, ray-finned fish and tetrapods; it would seem that the protein-coding portion of this gene family, along with several others (Supplementary Note 9, Supplementary Tables 12–16 and Supplementary Fig. 10), have remained largely conserved during the vertebrate land transition (Supplementary Fig. 11).

As vertebrates transitioned to a new land environment, changes occurred not only in gene content but also in the regulation of existing genes. Conserved non-coding elements (CNEs) are strong candidates for gene regulatory elements. They can act as promoters, enhancers, repressors and insulators^{32,33}, and have been implicated as major facilitators of evolutionary change³⁴. To identify CNEs that originated in the most recent common ancestor of tetrapods, we predicted CNEs that evolved in various bony vertebrate (that is, ray-finned fish, coelacanth and tetrapod) lineages and assigned them to their likely branch points of origin. To detect CNEs, conserved sequences in the human genome were identified using MULTIZ alignments of bony vertebrate genomes, and then known protein-coding sequences, untranslated regions (UTRs) and known RNA genes were excluded. Our analysis identified 44,200 ancestral tetrapod CNEs that originated after the divergence of the coelacanth lineage. They represent 6% of the 739,597 CNEs that are under constraint in the bony vertebrate lineage. We compared the ancestral tetrapod CNEs to mouse embryo ChIP-seq (chromatin immunoprecipitation followed by sequencing) data obtained using antibodies against p300, a transcriptional coactivator. This resulted in a sevenfold enrichment in the p300 binding sites for our candidate CNEs and confirmed that these CNEs are indeed enriched for gene regulatory elements.

Each tetrapod CNE was assigned to the gene whose transcription start site was closest, and gene-ontology category enrichment was calculated for those genes. The most enriched categories were involved with smell perception (for example, sensory perception of smell, detection of chemical stimulus and olfactory receptor activity). This is consistent with the notable expansion of olfactory receptor family genes in tetrapods compared with teleosts, and may reflect the necessity of a more tightly regulated, larger and more diverse repertoire of olfactory receptors for detecting airborne odorants as part of the terrestrial lifestyle. Other significant categories include morphogenesis (radial pattern formation, hind limb morphogenesis, kidney morphogenesis) and cell differentiation (endothelial cell fate commitment,

epithelial cell fate commitment), which is consistent with the body-plan changes required for land transition, as well as immunoglobulin VDJ recombination, which reflects the presumed response differences required to address the novel pathogens that vertebrates would encounter on land (Supplementary Note 10 and Supplementary Tables 17–24).

A major innovation of tetrapods is the evolution of limbs characterized by digits. The limb skeleton consists of a stylopod (humerus or femur), the zeugopod (radius and ulna, or tibia and fibula), and an autopod (wrist or ankle, and digits). There are two major hypotheses about the origins of the autopod; that it was a novel feature of tetrapods, and that it has antecedents in the fins of fish³⁵ (Supplementary Note 11 and Supplementary Fig. 12). We examine here the Hox regulation of limb development in ray-finned fish, coelacanth and tetrapods to address these hypotheses.

In mouse, late-phase digit enhancers are located in a gene desert that is proximal to the HOX-D cluster³⁶. Here we provide an alignment of the HOX-D centromeric gene desert of coelacanth with those of tetrapods and ray-finned fishes (Fig. 2a). Among the six *cis*-regulatory sequences previously identified in this gene desert³⁶, three sequences show sequence conservation restricted to tetrapods (Supplementary Fig. 13). However, one regulatory sequence (island 1) is shared by tetrapods and coelacanth, but not by ray-finned fish (Fig. 2b and Supplementary Fig. 14). When tested in a transient transgenic assay in mouse, the coelacanth sequence of island 1 was able to drive reporter expression in a limb-specific pattern (Fig. 2c). This suggests that island 1 was a lobe-fin developmental enhancer in the fish ancestor of tetrapods that was then coopted into the autopod enhancer of modern tetrapods. In this case, the autopod developmental regulation was derived from an ancestral lobe-finned fish regulatory element.

Changes in the urea cycle provide an illuminating example of the adaptations associated with transition to land. Excretion of nitrogen is a major physiological challenge for terrestrial vertebrates. In aquatic environments, the primary nitrogenous waste product is ammonia, which is readily diluted by surrounding water before it reaches toxic levels, but on land, less toxic substances such as urea or uric acid must be produced instead (Supplementary Fig. 15). The widespread and almost exclusive occurrence of urea excretion in amphibians, some turtles and mammals has led to the hypothesis that the use of urea as the main nitrogenous waste product was a key innovation in the vertebrate transition from water to land³⁷.

With the availability of gene sequences from coelacanth and lungfish, it became possible to test this hypothesis. We used a branch-site model

in the HYPHY package³⁸, which estimates the ratio of synonymous (dS) to non-synonymous (dN) substitutions (ω values) among different branches and among different sites (codons) across a multiple-species sequence alignment. For the rate-limiting enzyme of the hepatic urea cycle, carbamoyl phosphate synthase I (CPS1), only one branch of the tree shows a strong signature of selection ($P = 0.02$), namely the branch leading to tetrapods and the branch leading to amniotes (Fig. 3); no other enzymes in this cycle showed a signature of selection. Conversely, mitochondrial arginase (ARG2), which produces extrahepatic urea as a byproduct of arginine metabolism but is not involved in the production of urea for nitrogenous waste disposal, did not show any evidence of selection in vertebrates (Supplementary Fig. 16). This leads us to conclude that adaptive evolution occurred in the hepatic urea cycle during the vertebrate land transition. In addition, it is interesting to note that of the five amino acids of CPS1 that changed between coelacanth and tetrapods, three are in important domains (the two ATP-binding sites and the subunit interaction domain) and a fourth is known to cause a malfunctioning enzyme in human patients if mutated³⁹.

The adaptation to a terrestrial lifestyle necessitated major changes in the physiological environment of the developing embryo and fetus, resulting in the evolution and specialization of extra-embryonic membranes of the amniote mammals⁴⁰. In particular, the placenta is a complex structure that is critical for providing gas and nutrient exchange between mother and fetus, and is also a major site of haematopoiesis⁴¹.

We have identified a region of the coelacanth HOX-A cluster that may have been involved in the evolution of extra-embryonic structures in tetrapods, including the eutherian placenta. Global alignment of the coelacanth *Hoxa14*–*Hoxa13* region with the homologous regions of the horn shark, chicken, human and mouse revealed a CNE just upstream of the coelacanth *Hoxa14* gene (Supplementary Fig. 17a). This conserved stretch is not found in teleost fishes but is highly conserved among horn shark, chicken, human and mouse despite the fact that the chicken, human and mouse have no *Hoxa14* orthologues, and that the horn shark *Hoxa14* gene has become a pseudogene. This CNE, HA14E1, corresponds to the proximal promoter-enhancer region of the *Hoxa14* gene in *Latimeria*. HA14E1 is more than 99% identical between mouse, human and all other sequenced mammals, and would therefore be considered to be an ultra-conserved element⁴². The high level of conservation suggests that this element, which already possessed promoter activity, may have been coopted for other functions despite the loss of the *Hoxa14* gene in amniotes (Supplementary Fig. 17bc). Expression of human HA14E1 in a mouse

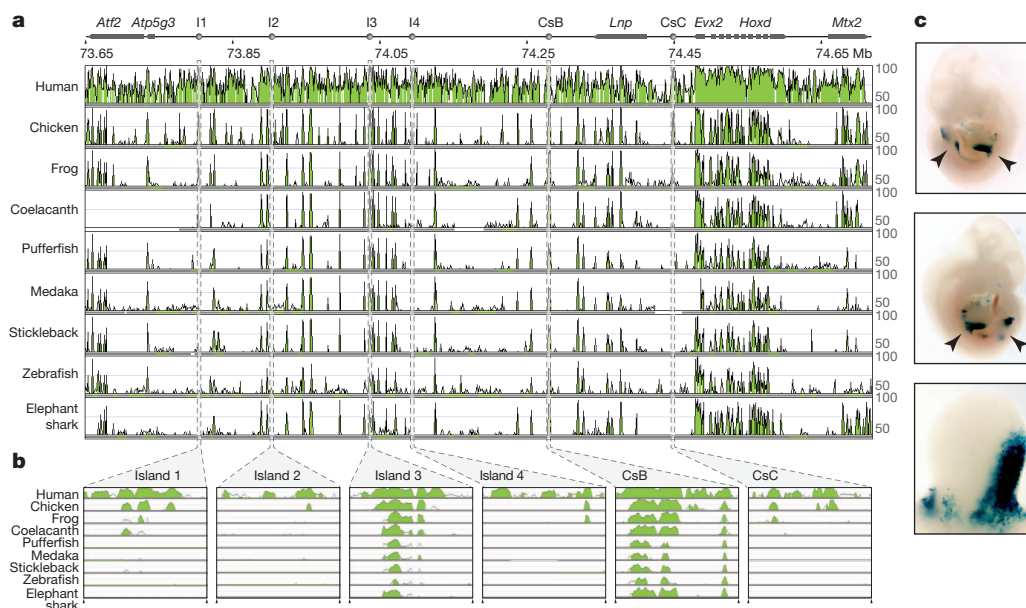


Figure 2 | Alignment of the HOX-D locus and an upstream gene desert identifies conserved limb enhancers. **a**, Organization of the mouse HOX-D locus and centromeric gene desert, flanked by the *Atf2* and *Mtx2* genes. Limb regulatory sequences (I1, I2, I3, I4, CsB and CsC) are noted. Using the mouse locus as a reference (NCBI and mouse genome sequencing consortium NCBI37/mm9 assembly), corresponding sequences from human, chicken, frog, coelacanth, pufferfish, medaka, stickleback, zebrafish and elephant shark were aligned. Alignment shows regions of homology between tetrapod, coelacanth and ray-finned fishes. **b**, Alignment of vertebrate *cis*-regulatory elements I1, I2, I3, I4, CsB and CsC. **c**, Expression patterns of coelacanth island I in a transgenic mouse. Limb buds are indicated by arrowheads in the first two panels. The third panel shows a close-up of a limb bud.

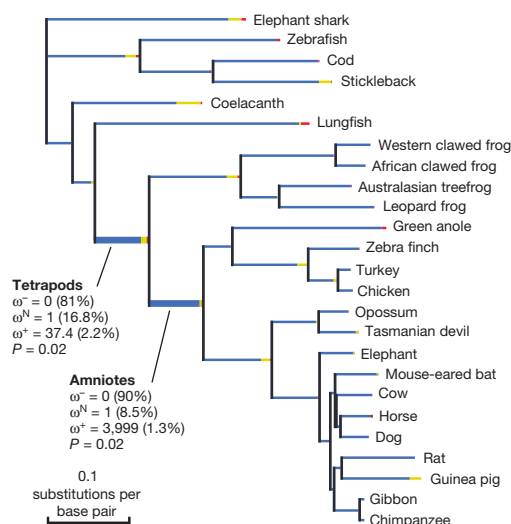


Figure 3 | Phylogeny of *Cps1* coding sequences is used to determine positive selection within the urea cycle. Branch lengths are scaled to the expected number of substitutions per nucleotide, and branch colours indicate the strength of selection (dN/dS or ω). Red, positive or diversifying selection ($\omega > 1$); blue, purifying selection ($\omega < 1$); yellow, neutral evolution ($\omega = 1$). Thick branches indicate statistical support for evolution under episodic diversifying selection. The proportion of each colour represents the fraction of the sequence undergoing the corresponding class of selection.

transient transgenic assay did not give notable expression in the embryo proper at day 11.5 (information is available online at the VISTA enhancer browser website; http://enhancer.lbl.gov/cgi-bin/imagdb3.pl?form=presentation&show=1&experiment_id=501&organism_id=1), which was unexpected as its location would predict that it would regulate axial structures caudally⁴³. A similar experiment in chick embryos using the chicken HA14E1 also showed no activity in the anteroposterior axis. However, strong expression was observed in the extraembryonic area vasculosa of the chick embryo (Fig. 4a). Examination of a *Latimeria* BAC *Hoxa14*-reporter transgene in mouse embryos showed that the *Hoxa14* gene is specifically expressed in a subset of cells in an extra-embryonic region at embryonic day 8.5 (Fig. 4b).

These findings suggest that the HA14E1 region may have been evolutionarily recruited to coordinate regulation of posterior HOX-A genes (*Hoxa13*, *Hoxa11* and *Hoxa10*), which are known to be expressed in the mouse allantois and are critical for early formation of the mammalian placenta⁴⁴. Although *Latimeria* does not possess a placenta, it gives birth to live young and has very large, vascularised eggs, but the relationship between *Hoxa14*, the HA14E1 enhancer and blood island formation in the coelacanth remains unknown.

The coelacanth lacks immunoglobulin-M

Immunoglobulin-M (IgM), a class of antibodies, has been reported in all vertebrate species that have been characterized so far, and is considered to be indispensable for adaptive immunity⁴⁵. Interestingly, IgM genes cannot be found in coelacanth, despite an exhaustive search of the coelacanth sequence data, and even though all other major components of the immune system are present (Supplementary Note 12 and Supplementary Fig. 18). Instead, we found two IgW genes (Supplementary Figs 19–21); immunoglobulin genes that are found only in lungfish and cartilaginous fish and are believed to have originated in the ancestor of jawed vertebrates⁴⁶ but subsequently lost in teleosts and tetrapods. IgM was similarly absent from the *Latimeria* RNA-seq data, although both IgW genes were found as transcripts. To characterize further the apparent absence of IgM, we screened large genomic *L. menadoensis* libraries exhaustively using a number of strategies and probes. We also carried out polymerase chain reaction (PCR) with degenerate primers that should universally amplify IgM sequences. The lack of IgM in *Latimeria* raises questions as to how coelacanth

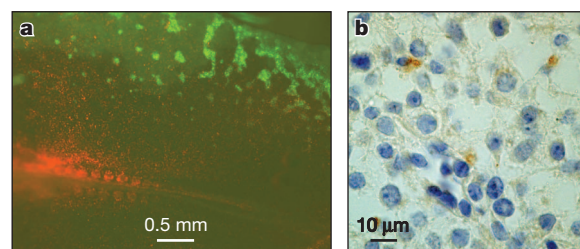


Figure 4 | Transgenic analysis implicates involvement of *Hox* CNE HA14E1 in extraembryonic activities in the chick and mouse. **a**, Chicken HA14E1 drives reporter expression in blood islands in chick embryos. A construct containing chicken HA14E1 upstream of a minimal (thymidine kinase) promoter driving enhanced green fluorescent protein (eGFP) was electroporated in HH4-stage chick embryos together with a nuclear mCherry construct. GFP expression was analysed at stage approximately HH11. The green aggregations and punctate staining are observed in the blood islands and developing vasculature. **b**, Expression of *Latimeria Hoxa14*-reporter transgene in the developing placental labyrinth of a mouse embryo. A field of cells from the labyrinth region of an embryo at embryonic day 8.5 from a BAC transgenic line containing coelacanth *Hoxa9–Hoxa14* (ref. 49) in which the *Hoxa14* gene had been supplanted with the gene for red fluorescence protein (RFP). Immunohistochemistry was used to detect RFP (brown staining in a small number of cells).

B cells respond to microbial pathogens and whether the IgW molecules can serve a compensatory function, even though there is no indication that the coelacanth IgW was derived from vertebrate IgM genes.

Discussion

Since its discovery, the coelacanth has been referred to as a 'living fossil', owing to its morphological similarities to its fossil ancestors¹. However, questions have remained as to whether it is indeed evolving slowly, as morphological stasis does not necessarily imply genomic stasis. In this study, we have confirmed that the protein-coding genes of *L. chalumnae* show a decreased substitution rate compared to those of other sequenced vertebrates, even though its genome as a whole does not show evidence of low genome plasticity. The reason for this lower substitution rate is still unknown, although a static habitat and a lack of predation over evolutionary timescales could be contributing factors to a lower need for adaptation. A closer examination of gene families that show either unusually high or low levels of directional selection indicative of adaptation in the coelacanth may provide information on which selective pressures acted, and which pressures did not act, to shape this evolutionary relict (Supplementary Note 13 and Supplementary Fig. 22).

The vertebrate land transition is one of the most important steps in our evolutionary history. We conclude that the closest living fish to the tetrapod ancestor is the lungfish, not the coelacanth. However, the coelacanth is critical to our understanding of this transition, as the lungfish have intractable genome sizes (estimated at 50–100 Gb)⁴⁷. Here we have examined vertebrate adaptation to land through coelacanth whole-genome analysis, and have shown the potential of focused analysis of specific gene families involved in this process. Further study of these changes between tetrapods and the coelacanth may provide important insights into how a complex organism like a vertebrate can markedly change its way of life.

METHODS SUMMARY

A full description of methods, including information on sample collection, sequencing, assembly, annotation, all sequence analysis and functional validation, can be found in the Supplementary Information.

Received 5 September 2012; accepted 20 February 2013.

- Smith, J. L. B. A living fish of mesozoic type. *Nature* **143**, 455–456 (1939).
- Nulens, R., Scott, L. & Herbin, M. *An Updated Inventory of All Known Specimens of the Coelacanth*, *Latimeria* Spp. *Smithiana* Vol. 3 (South African Institute for Aquatic Biodiversity, 2010).

3. Erdmann, M. V., Caldwell, R. L. & Kasim Moosa, M. Indonesian 'king of the sea' discovered. *Nature* **395**, 335 (1998).
4. Smith, J. L. B. *Old Fourlegs: the Story of the Coelacanth* (Longmans, Green, 1956).
5. Zhu, M. et al. Earliest known coelacanth skull extends the range of anatomically modern coelacanths to the Early Devonian. *Nature Commun.* **3**, 772 (2012).
6. Zimmer, C. *At the Water's Edge: Fish with Fingers, Whales with Legs, and How Life Came Ashore but then Went Back to Sea* (Free Press, 1999).
7. Zardoya, R. & Meyer, A. The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). *Genetics* **146**, 995–1010 (1997).
8. Amemiya, C. T. et al. Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc. Natl Acad. Sci. USA* **107**, 3622–3627.
9. Larsson, T. A., Larson, E. T. & Larhammar, D. Cloning and sequence analysis of the neuropeptide Y receptors Y5 and Y6 in the coelacanth *Latimeria chalumnae*. *Gen. Comp. Endocrinol.* **150**, 337–342 (2007).
10. Noonan, J. P. et al. Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. *Genome Res.* **14**, 2397–2405 (2004).
11. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl Acad. Sci. USA* **108**, 1513–1518 (2011).
12. Bogart, J. P., Balon, E. K. & Bruton, M. N. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. *J. Hered.* **85**, 322–325 (1994).
13. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
14. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotech.* **29**, 644–652 (2011).
15. Pallavicini, A. et al. Analysis of the transcriptome of the Indonesian coelacanth *Latimeria menadoensis*. *BMC Genomics* (in the press).
16. Schultze, H. P. & Trueb, L. *Origins of the Higher Groups of Tetrapods: Controversy and Consensus*. (Cormstock Publishing Associates, 1991).
17. Meyer, A. & Dolven, S. I. Molecules, fossils, and the origin of tetrapods. *J. Mol. Evol.* **35**, 102–113 (1992).
18. Brinkmann, H., Venkatesh, B., Brenner, S. & Meyer, A. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc. Natl Acad. Sci. USA* **101**, 4900–4905 (2004).
19. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
20. Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).
21. Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607 (1993).
22. Bejerano, G. et al. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87–90 (2006).
23. Voss, S. R. et al. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* **21**, 1306–1312 (2011).
24. Smith, J. J. & Voss, S. R. Gene order data from a model amphibian (*Ambystoma*): new perspectives on vertebrate genome structure and evolution. *BMC Genomics* **7**, 219 (2006).
25. Inoue, J. G., Miya, M., Venkatesh, B. & Nishida, M. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* **349**, 227–235 (2005).
26. Holder, M. T., Erdmann, M. V., Wilcox, T. P., Caldwell, R. L. & Hillis, D. M. Two living species of coelacanths? *Proc. Natl Acad. Sci. USA* **96**, 12616–12620 (1999).
27. Canapa, A. et al. Composition and phylogenetic analysis of vitellogenin coding sequences in the Indonesian coelacanth *Latimeria menadoensis*. *J. Exp. Zool.* **B318**, 404–416 (2012).
28. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
29. Zhang, J. et al. Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* **466**, 234–237 (2010).
30. Jovel, R. et al. Evolution of developmental regulation in the vertebrate FgfD subfamily. *J. Exp. Zool.* **B314**, 33–56 (2010).
31. Braasch, I. & Postlethwait, J. H. The teleost agouti-related protein 2 gene is an ohnolog gone missing from the tetrapod genome. *Proc. Natl Acad. Sci. USA* **108**, E47–E48 (2011).
32. Navratilova, P. et al. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev. Biol.* **327**, 526–540 (2009).
33. Xie, X. et al. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl Acad. Sci. USA* **104**, 7145–7150 (2007).
34. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
35. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
36. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
37. Wright, P. A. Nitrogen excretion: three end products, many physiological roles. *J. Exp. Biol.* **198**, 273–281 (1995).
38. Kosakovsky Pond, S. L. et al. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* **28**, 3033–3043 (2011).
39. Häberle, J. et al. Molecular defects in human carbamoyl phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations. *Hum. Mutat.* **32**, 579–589 (2011).
40. Carroll, R. L. *Vertebrate Paleontology and Evolution* (W.H. Freeman and Company, 1988).
41. Gekas, C. et al. Hematopoietic stem cell development in the placenta. *Int. J. Dev. Biol.* **54**, 1089–1098 (2010).
42. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
43. Wellik, D. M. Hox patterning of the vertebrate axial skeleton. *Dev. Dyn.* **236**, 2454–2463 (2007).
44. Scotti, M. & Kmita, M. Recruitment of 5' Hoxa genes in the allantois is essential for proper extra-embryonic function in placental mammals. *Development* **139**, 731–730 (2012).
45. Bengtén, E. et al. Immunoglobulin isotypes: structure, function, and genetics. *Curr. Top. Microbiol. Immunol.* **248**, 189–219 (2000).
46. Ota, T., Rast, J. P., Litman, G. W. & Amemiya, C. T. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. *Proc. Natl Acad. Sci. USA* **100**, 2501–2506 (2003).
47. Gregory, T. R. *The Evolution of the Genome* 1–71 (Elsevier Academic, 2004).
48. Stamatakis, A., Ludwig, T. & Meier, H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
49. Smith, J. J., Sumiyama, K. & Amemiya, C. T. A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome. *Mol. Biol. Evol.* **29**, 985–993 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements Acquisition and storage of *Latimeria chalumnae* samples was supported by grants from the African Coelacanth Ecosystem Programme of the South African National Department of Science and Technology. Generation of the *Latimeria chalumnae* and *Protopterus annectens* sequences by the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard University was supported by grants from the National Human Genome Research Institute (NHGRI). K.L.T. is the recipient of a EURYL award from the European Science Foundation. We would also like to thank the Genomics Sequencing Platform of the Broad Institute for sequencing the *L. chalumnae* genome and *L. chalumnae* and *P. annectens* transcriptomes, S. Ahmada, R. Stobbs and the Association pour le Protection de Gombesa (APG) for their help in obtaining coelacanth samples, Y. Zhao for the use of data from *Rana chensinensis*, and L. Gaffney, C. Hamilton and J. Westlund for assistance with figure preparation.

Author Contributions J.A., C.T.A., A.M. and K.L.T. planned and oversaw the project. R.A.D. and C.T.A. provided blood and tissues for sequencing. C.T.A. and M.L. prepared the DNA for sequencing. I.M., S.G., D.P., F.J.R., T.S. and D.B.J. assembled the genome. N.R.S. and C.T.A. prepared RNA from *L. chalumnae* and *P. annectens*, and L.F. and J.Z.L. made the *L. chalumnae* RNA-seq library. A.C., M.B., M.A.B., M.F., F.B., G.S., A.M.F., A.P., M.G., G.D.M., J.T.-M. and E.O. sequenced and analysed the *L. menadoensis* RNA-seq library. B.A., S.M.J.S., S.W., M.S.C. and M.Y. annotated the genome. W.H. and C.P.P. carried out the annotation and analysis of long non-coding RNAs. P.F.S., S.H., A.N., H.T. and S.J.P. annotated non-coding RNAs. M.G., G.D.M., A.P., M.R. and C.T.A. compared *L. chalumnae* and *L. menadoensis* sequences. H.B., D.B. and H.P. carried out the phylogenomic analysis. T.Ma. and A.M. performed the gene relative-rate analysis. A.C., J.G., S.P., B.P., P.v.H. and U.H. carried out the analysis, annotation and statistical enrichment of *L. chalumnae* specific gene duplications. N.F. and A.M. analysed the homeobox gene repertoires. G.L.B. and A.L.E. analysed the chaperone genes. D.C., S.F., O.S., J.-N.V., M.S. and A.M. analysed transposable elements. J.J.S. analysed large-scale rearrangements in vertebrate genomes. I.B., J.H.P., N.F. and S.K. analysed genes lost in tetrapods. T.Mi. analysed actinodin and pectoral-fin musculature. C.O. and M.S. analysed selection in urea cycle genes. A.P.L. and B.V. carried out the conserved non-coding element analysis. I.S., N.R., V.R., N.S. and C.J.T. carried out the analysis of autopodial CNEs. K.S., T.S.-S. and C.T.A. examined the evolution of a placenta-related CNE. N.R.S., G.W.L., M.G.M., T.O. and C.T.A. performed the IgM analysis. J.A., C.T.A., A.M. and K.L.T. wrote the paper with input from other authors.

Author Information Genome assemblies, transcriptomes and mitochondrial DNA sequences have been deposited in GenBank/EMBL/DBJ. The *L. chalumnae* genome assembly has been deposited under the accession number AFYH00000000. The *L. chalumnae* transcriptome has been deposited under the accession number SRX117503 and the *P. annectens* transcriptomes have been deposited under the accession numbers SRX152529, SRX152530 and SRX152531. The *P. annectans* mitochondrial DNA sequence was deposited under the accession number JX568887. All animal experiments were approved by the MIT Committee for Animal Care. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T.A. (camemiya@benaroyaresearch.org), J.A. (jalfoldi@broadinstitute.org), A.M. (axel.meyer@uni-konstanz.de) or K.L.T. (kersli@broadinstitute.org).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Gating of the TrkH ion channel by its associated RCK protein TrkA

Yu Cao^{1*}, Yaping Pan^{1,2*}, Hua Huang^{1,2*}, Xiangshu Jin³, Elena J. Levin^{1,2}, Brian Kloss⁴ & Ming Zhou^{1,2}

TrkH belongs to a superfamily of K⁺ transport proteins required for growth of bacteria in low external K⁺ concentrations. The crystal structure of TrkH from *Vibrio parahaemolyticus* showed that TrkH resembles a K⁺ channel and may have a gating mechanism substantially different from K⁺ channels. TrkH assembles with TrkA, a cytosolic protein comprising two RCK (regulate the conductance of K⁺) domains, which are found in certain K⁺ channels and control their gating. However, fundamental questions on whether TrkH is an ion channel and how it is regulated by TrkA remain unresolved. Here we show single-channel activity of TrkH that is upregulated by ATP via TrkA. We report two structures of the tetrameric TrkA ring, one in complex with TrkH and one in isolation, in which the ring assumes two markedly different conformations. These results suggest a mechanism for how ATP increases TrkH activity by inducing conformational changes in TrkA.

K⁺ is concentrated in all living cells and is essential for many physiological processes. In bacteria, homeostasis of K⁺ is largely mediated by specialized K⁺ transport proteins known as the superfamily of K⁺ transport (SKT) proteins¹. The bacterial TrkH, TrkG and KtrB proteins form the largest subfamily of SKT proteins, and the crystal structure of TrkH from *Vibrio parahaemolyticus*, VpTrkH, was reported recently². Each of the four connected homologous domains of TrkH resembles a subunit of the KcsA K⁺ channel, comprising two transmembrane helices connected by a re-entrant P-loop (M1-P-M2)³, and the four domains encircle a central ion permeation pathway. In addition, TrkH has two structural features atypical of K⁺ channels: a dimeric quaternary structure, with each protomer containing its own pore; and a long membrane-embedded loop, inserted into the middle of a pore-lining helix, that occludes the ion permeation pathway. The loop is conserved across the TrkH/TrkG/KtrB family, and previous studies have shown that deletion of this intramembrane loop increases KtrB-mediated K⁺ uptake, indicating that it may act as a gate^{4,5}.

The TrkH, TrkG and KtrB integral membrane proteins form a complex with an intracellular regulatory protein, and the genes encoding the membrane protein and the regulatory protein reside in the same operon in many bacteria. The protein that assembles with TrkH and TrkG is TrkA, whereas KtrB interacts with KtrA^{6–8}. Both TrkA and KtrA are composed of RCK domains^{9–16}, which can be classified into two types based on their cognate ligands: bacterial RCK domains possessing conserved GXGXXG motifs that bind to nucleotides^{8,12,14,17,18}, sometimes referred to as KTN domains, and the bacterial and mammalian RCKs that lack a nucleotide-binding motif and bind to ions such as Ca²⁺ (refs 11, 15, 16, 19). The organization of RCK domains varies considerably. For example, the KtrA and TrkA protomers contain one and two RCK domains, respectively, expressed as separate cytoplasmic proteins, whereas for K⁺ channels the RCK domains are directly connected to the termini of their associated channels. Previous RCK structures suggest that regardless of gene architecture, a total of eight RCK domains associate in a configuration known as a gating ring^{9,11,14–16,19}.

The TrkH, TrkG and KtrB proteins were initially thought to be either H⁺- or Na⁺-dependent K⁺ transporters^{20–22}, but their sequence and structural similarity to K⁺ channels suggests that these proteins could function as ion channels. However, ion channel activity has never been demonstrated for any member of the TrkH/TrkG/KtrB family of proteins. Furthermore, if TrkH is an ion channel, what effect does its associated protein TrkA have on gating of the channel? Current models of gating in K⁺ channels with RCK domains propose that a dilation in the diameter of the gating ring directly translates into movement of the four pore-lining helices and opening of the permeation pathway^{9,11,15,16,19,23}. However, this model of a four-fold symmetric gating ring expansion seems to be incompatible with the dimeric architecture of TrkH, and does not account for the intramembrane loop. In this study, we attempt to address these unresolved issues by characterizing the structure and function of the TrkH–TrkA complex.

Single-channel activities of the TrkH–TrkA complex

Patch clamp was performed on *Escherichia coli* spheroplasts heterologously expressing TrkH and TrkA from *V. parahaemolyticus* using an inside-out configuration. A single-channel current trace in 200 mM symmetrical K⁺ solution is shown in Fig. 1a. Occasional bursts of channel activities were observed, and these bursts were interrupted by prolonged closures. Within each burst of openings there are two current levels: level 1 and 2, with level 2 approximately double the size of level 1 in amplitude (Supplementary Table 1). Every burst of openings starts with a sojourn from the baseline directly to level 2 followed by openings alternating between levels 1 and 2, and terminates >91% of the time directly from level 2. This behaviour is inconsistent with two entirely independent channels, but strongly suggests a single TrkH dimer, the two pores of which open and close in a mostly concerted fashion.

As a control, currents from spheroplasts expressing only TrkH were also recorded. Similar bursts of channel activities with two levels of amplitudes were observed (Supplementary Table 1 and Supplementary Fig. 1a). However, the overall open probability of TrkH without TrkA is higher than that of the complex (0.65 ± 0.03 versus 0.17 ± 0.01), suggesting that TrkA affects the activity of TrkH. To verify that the

¹Department of Physiology & Cellular Biophysics, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ²Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ³National Institute of Biological Sciences, Beijing 102206, China. ⁴New York Consortium on Membrane Protein Structure, New York Structural Biology Center, New York, New York 10027, USA.

*These authors contributed equally to this work.

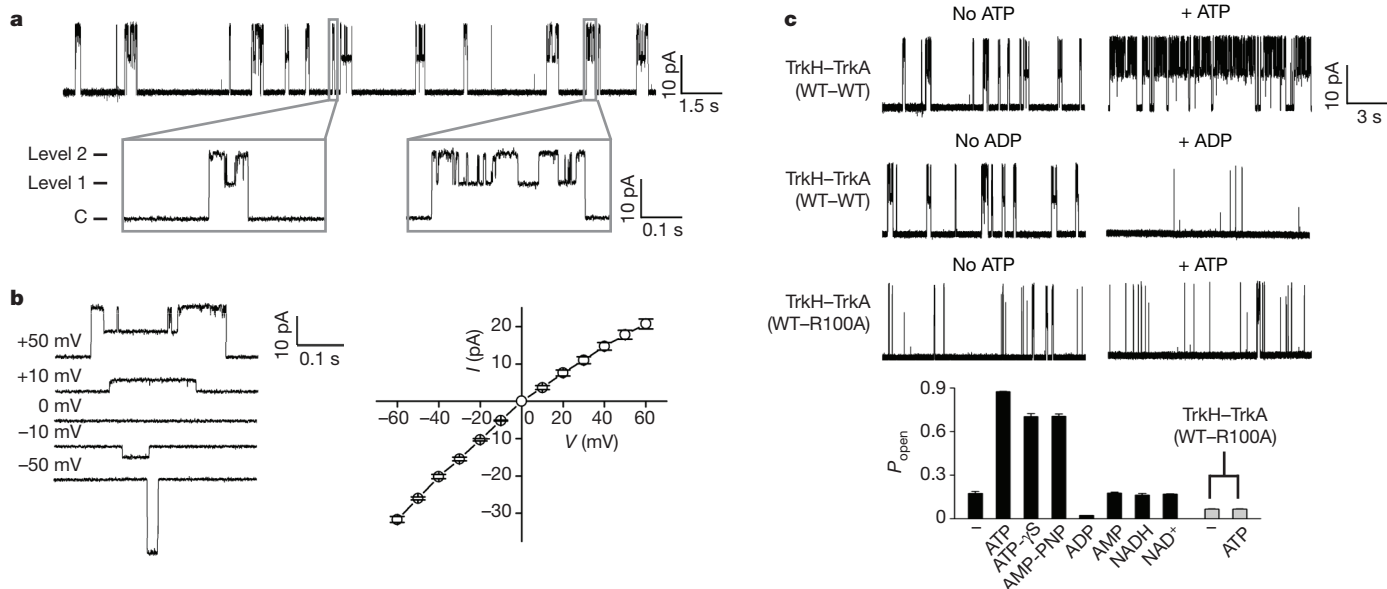


Figure 1 | Single-channel activities of the TrkH-TrkA complex. **a**, Single-channel currents through the TrkH-TrkA complex were recorded in symmetrical 200 mM KCl. The current trace is shown at a holding potential of +50 mV. The closed state is indicated as 'C'. The grey rectangles show two individual bursts of channel openings. **b**, The open-channel current-voltage relationship. Current traces recorded at different voltages are shown in the left panel, and current amplitudes are plotted in the right panel. **c**, Current traces of

single-channel currents were from TrkH, a cysteine substitution was made to a pore-lining residue (I220C) based on the known TrkH structure². Perfusion of the cysteine-modifying reagent (2-(trimethylammonium)ethyl) methanethiosulphonate (MTSET)²⁴ reversibly reduced current amplitudes as well as the overall open probability for the I220C mutant for either TrkH alone or the TrkH-TrkA complex (Supplementary Fig. 1b, c). Wild-type TrkH was not affected by MTSET (Supplementary Fig. 1c).

The open-channel current-voltage (I - V plot) relationship is plotted in Fig. 1b and shows slight inward rectification: level 2 has a chord conductance of 350 pS at +60 mV and 550 pS at -60 mV. Selectivity of the TrkH ion channel was examined by changing the solution on the intracellular side of the patch. Although TrkH does not discriminate Rb^{+} and Cs^{+} from K^{+} , it conducts Na^{+} and Li^{+} at a significantly slower rate (Supplementary Table 1 and Supplementary Fig. 2). Given that the K^{+} channel signature sequence, TVGYG, which lines the selectivity filter, is not highly conserved in TrkH, the weak selectivity of TrkH is not surprising. In addition, substituting most of the intracellular Cl^{-} with a large organic anion, gluconate, does not affect the reversal potential, indicating that the TrkH channel does not conduct Cl^{-} (Supplementary Fig. 2d).

Because TrkA contains a conserved nucleotide-binding motif, we examined channel activity in the presence of NAD^{+} , NADH and ATP. Whereas perfusion of NADs to the intracellular side of the channel did not produce significant changes in the currents, perfusion of 5 mM ATP substantially increased the open probability (P_{open}) of the TrkH-TrkA complex (Fig. 1c and Supplementary Fig. 3). The increase in P_{open} is the result of both a reduced time between bursts and a prolonged burst duration (Fig. 1c). To verify that this effect was mediated by TrkA, we mutated R100 on TrkA, which forms part of the ATP-binding site (see 'Structure of the isolated TrkA tetramer' below), to alanine, and the mutant no longer responds to ATP with an increase in P_{open} (Fig. 1c). This increase in P_{open} does not require ATP hydrolysis, because two non-hydrolysable ATP analogues, adenosine 5'-O-(3-thiotriphosphate) (ATP- γ S) and adenosine 5'-(β , γ -imidotriphosphate) (AMP-PNP), both produced a similar increase

the TrkH-TrkA (wild type (WT)-WT) complex and the TrkH-TrkA (WT-R100A) complex in the absence (left) and presence (right) of 5 mM ATP or ADP. The P_{open} after perfusion of different ligands is plotted in the bar graph. Added ligand concentrations are as follows: ATP, 5 mM; ATP- γ S, 1 mM; AMP-PNP, 1 mM; ADP, 5 mM; AMP, 5 mM; NADH, 1 mM; NAD^{+} , 1 mM. Error bars are s.e.m. of 3-9 independent measurements.

in P_{open} (Fig. 1c). This result is consistent with earlier studies indicating that the Trk system in bacteria requires ATP²⁵ but is not an ATPase²⁶. Notably, ADP had the opposite effect to that of ATP and significantly reduced the P_{open} , whereas AMP had a negligible effect (Fig. 1c and Supplementary Fig. 3).

Structure of the TrkH-TrkA complex

To understand how TrkH and TrkA interact, we co-expressed the two proteins and crystallized the stable purified complex (Supplementary Fig. 4). Although NADH does not affect channel activity (Fig. 1c), inclusion of NADH in the crystallization solution improved crystal size and quality. The structure of the complex was solved at a resolution of 3.8 Å (Supplementary Fig. 5 and Supplementary Table 2). In the asymmetric unit, two TrkH dimers interact with a ring of four TrkA protomers; this probably occurs because the internal symmetry of the TrkA tetramer creates two equivalent interfaces on both sides of the ring, as discussed in detail below. In the membrane, one TrkH dimer would interact with one TrkA tetramer (Supplementary Fig. 6).

The tetrameric TrkA gating ring associates with the cytoplasmic face of the TrkH dimer (Fig. 2a, b). The structure of TrkH in the complex is almost identical to the previously solved structure of TrkH in isolation (Supplementary Fig. 7a). In particular, the pore of the ion conduction pathway is still obstructed by the intramembrane loop (Fig. 2b inset and Supplementary Fig. 7b). Each TrkH monomer interacts with TrkA through a cytosolic loop following the second half of the broken M2 helix in domain 3, D3M2b (Fig. 2b and Supplementary Fig. 4), which is tilted to lie nearly parallel to the bilayer and is connected directly to the intramembrane loop. The interface includes a number of salt bridges between positively charged residues on the D3M2b-D4 loop and negatively charged residues on TrkA (Supplementary Fig. 8b). A number of residues on the cytosolic side of TrkH are missing from the electron density, including an 18-residue stretch on the loop connecting domains 1 and 2 (Supplementary Fig. 8a). Although this cytosolic loop is not highly conserved among the TrkG, TrkH and KtrB proteins, because of its proximity to TrkA it may be involved in additional contacts not resolved in the crystal structure.

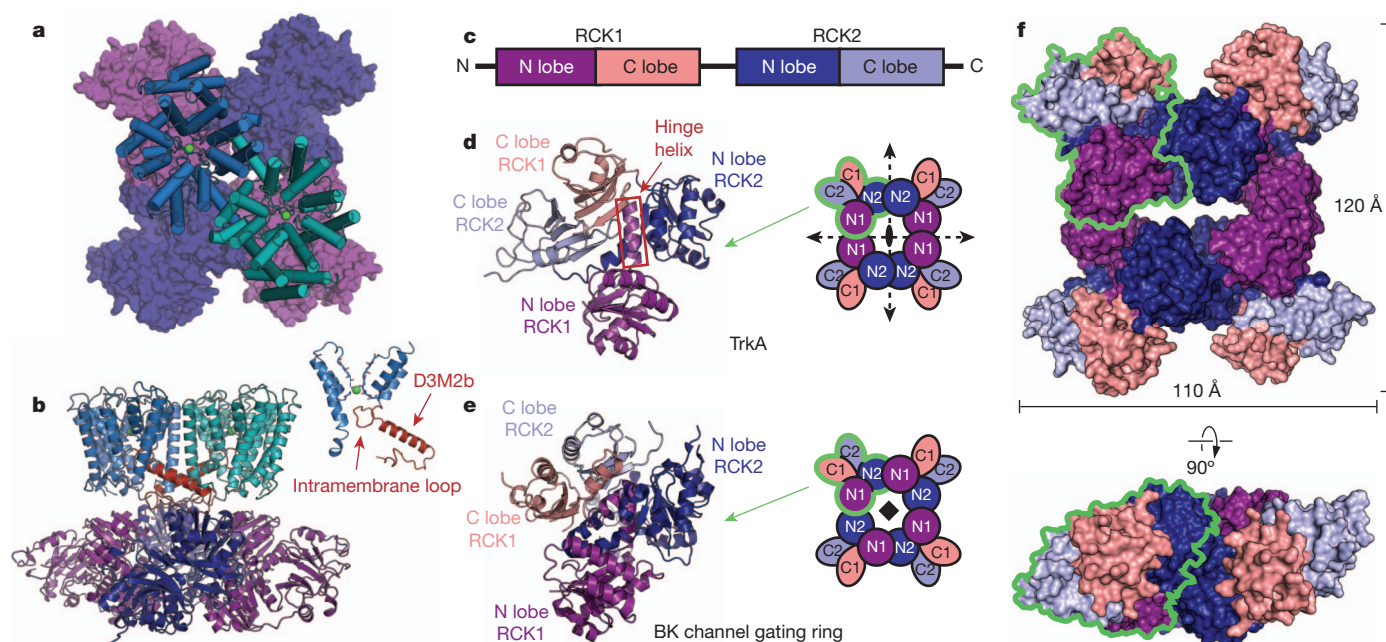


Figure 2 | Structure of the TrkH–TrkA complex. **a**, Structure of the complex between the TrkH dimer (light blue and teal) and the TrkA tetramer (dark blue and purple), viewed from the periplasmic side. The TrkA protomer is shown as a surface representation. **b**, The TrkH–TrkA complex viewed from within the plane of the membrane. Helix D3M2b in each TrkH subunit is highlighted in red. The inset on the right shows homologous domains 1 and 3 of TrkH, with the intramembrane loop and helix D3M2b highlighted. **c**, Schematic showing the organization of RCK domains in a TrkA subunit. **d**, **e**, Cartoon representation

of a TrkA protomer (**d**) or a protomer from the BK channel gating ring (**e**), coloured by subdomain according to the colour scheme in panel **c**. Schematics illustrating the organization of the tetrameric gating ring with the two-fold or four-fold symmetry axes marked are shown on the right. The green outlines delineate an individual subunit in each tetramer. **f**, Surface representation of the TrkA tetramer in the TrkH–TrkA complex, viewed from the periplasmic side (top) or parallel to the membrane (bottom). The green outline delineates an individual subunit.

Fold and quaternary structure of TrkA RCK gating ring

The TrkA protomer contains two tandem RCK domains, RCK1 and RCK2, so that the gating ring contains a total of eight RCK domains. Each RCK domain comprises an amino-terminal lobe assuming a Rossman fold, connected to a smaller carboxy-terminal lobe by a 'hinge' helix (Fig. 2c, d). A key difference between the TrkA structure and the BK and MthK channel-associated RCK structures^{11,15,16} is that the TrkA tetramer has D_2 point group symmetry, involving three perpendicular, intersecting, two-fold rotational symmetry axes (Fig. 2d, f), whereas the BK gating ring has C_4 symmetry (Fig. 2e). As a consequence of its 'dimer of dimers' arrangement, the TrkA tetramer provides a two-fold symmetric interaction surface to match the two-fold symmetry of the TrkH dimer. Additionally, the tetramer contains two distinct types of subunit interfaces: N1–N1 interfaces, formed between the N-terminal lobes of the RCK1 domains; and N2–N2 interfaces, formed by the N-terminal lobes of the RCK2 domains. In contrast, the four-fold symmetric BK gating ring contains four identical N1–N2 interfaces.

The TrkA protomer contains two GXGXXG nucleotide-binding motifs²⁷, one in the N-lobe of each RCK domain. The N2 lobe binding site in each TrkA protomer is occupied by a large electron density consistent with NADH, which was included as an additive to facilitate crystallization. At this low resolution the oxidized and reduced forms cannot be distinguished (Supplementary Fig. 9). The N1 lobe binding site is unoccupied.

Structure of the isolated TrkA tetramer

We were also able to obtain crystals of TrkA without TrkH and in the presence of 10 mM ATP- γ S that diffracted to a resolution of 3.05 Å, and solved the structure by molecular replacement (Fig. 3a and Supplementary Table 2). Isolated TrkA forms a tetramer in solution (Supplementary Fig. 10) as well as in the crystal lattice (Supplementary Fig. 11). Each protomer contains two large densities consistent with ATP- γ S molecules, with one bound to the GXGXXG motif in the

N lobe of each RCK domain, so that the gating ring contains a total of eight bound nucleotides (Supplementary Fig. 12).

Comparison of the structure of isolated TrkA to the structure from the TrkH–TrkA complex shows a substantial conformational change (Fig. 3a). This change is generated by markedly different behaviour of the N1–N1 and N2–N2 interfaces. The N lobes of the N2–N2 interface rotate and slide past each other (Fig. 3b and Supplementary Video 1), with one N2–N2 interface undergoing a larger translation than the other (Supplementary Fig. 13 and Supplementary Video 2). In contrast, the N1–N1 interface shows no change between the two structures (Fig. 3c and Supplementary Video 3). The different behaviour of the N1–N1 and N2–N2 interfaces converts TrkA from a roughly planar conformation to a twisted conformation (Fig. 3a).

The conformational change within individual protomers driving the 'flat' to 'twisted' transition in the gating ring can be visualized by aligning protomers from the ATP- γ S-bound structure and the complex by their immobile N1 lobes only, which reveals rigid-body rotations of the N2, C1 and C2 lobes about the hinge helix connecting N1 to C1 (Fig. 3d and Supplementary Video 4). Given that the two structures were obtained at low resolution and bound to non-physiological ligands, it is difficult to identify specific protein–ligand interactions that may drive this conformational change. However, we note that the γ -phosphate groups on the bound ATP- γ S molecules are located very close to the cleft between the two N lobes (Fig. 3e and Supplementary Fig. 13). The width of this cleft changes between the two structures, as measured by a change in the intersection angle of the hinge helices from 120° in the TrkH–TrkA structure to approximately 100° in the ATP- γ S-bound TrkA structure (Fig. 3e).

The change in the angle of intersection between hinge helices (Fig. 3e) is not unique to TrkA; similar movements have been observed in the KefC^{12,13}, KtrA¹⁴ and MthK¹⁹ gating rings. In contrast, the interfaces corresponding to the intersubunit interfaces in TrkA (referred to as 'assembly' interfaces in MthK and BK channels) seem to vary in their behaviour between the different gating rings. In BK

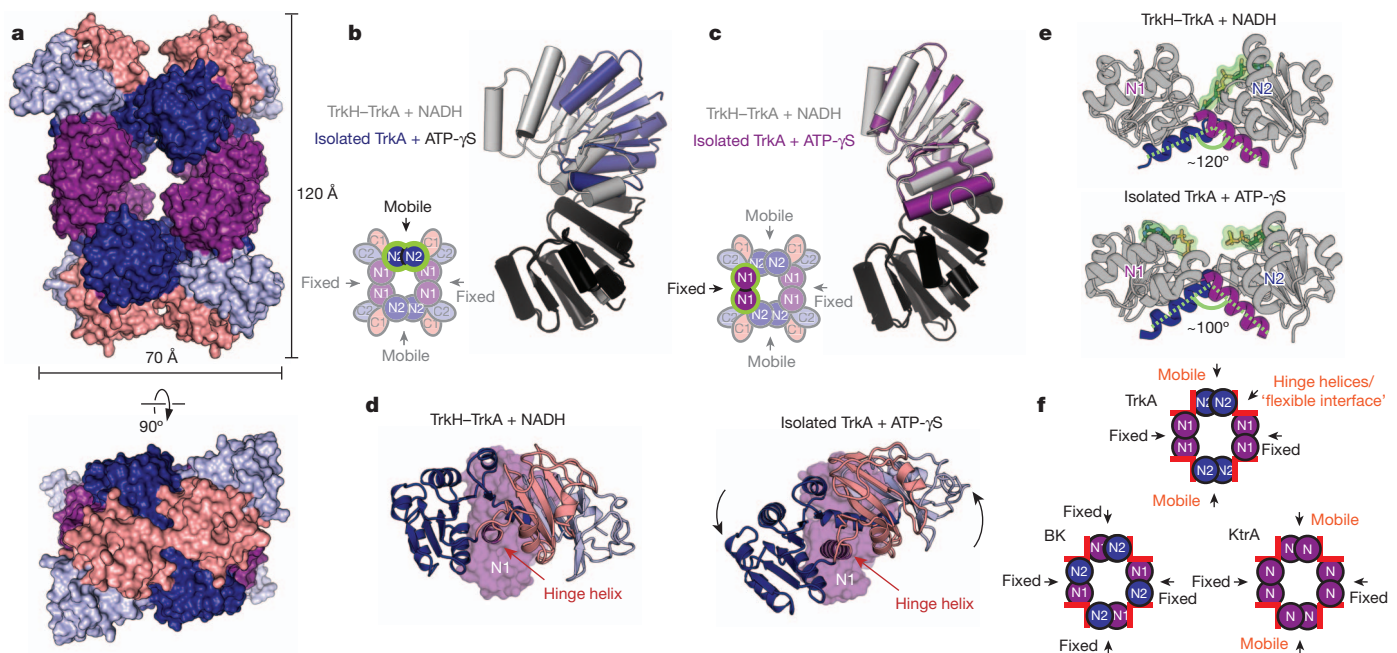


Figure 3 | Comparison of TrkA in the TrkA-only and TrkH-TrkA complex structures. **a**, Surface representation of the isolated TrkA tetramer, viewed from the periplasmic side (top) or parallel to the membrane (bottom), coloured by subdomain according to the colour scheme in Fig. 2c. **b**, **c**, Superposition of the N2–N2 interface (**b**) and N1–N1 interface (**c**) in the TrkH–TrkA (grey) and TrkA-only (blue or purple) structures. The alignment was based on the domains coloured black in both structures. **d**, A single TrkA protomer from the TrkH–TrkA complex structure (left) and a protomer from the TrkA-only structure (right) viewed from the same orientation after superposition by their

N1 subdomains. N1 is shown as surface representation, whereas the other three lobes are shown as cartoon representation. **e**, The N1 and N2 lobes in a subunit of TrkA from the NADH-bound TrkH–TrkA complex (top) and the ATP-γS-bound TrkA (bottom) structures are shown with their hinge helices highlighted. NADH or ATP-γS are shown in green. **f**, Cartoons of the TrkA, BK and KtrA gating rings showing only the N lobes. The hinge helices are marked with red rectangles. Description of interfaces as ‘fixed’ or ‘rotating’ is based on comparison of structures 2HMS and 2HMW (Protein Data Bank accessions) for TrkA, and 3NAF and 3U6N (Protein Data Bank accessions) for BK.

and MthK channels, all four assembly interfaces are fixed, forcing the ring to accommodate the change in the angle of the N-lobe cleft through a uniform, four-fold symmetric dilation. In KtrA and TrkA, two of the four assembly interfaces are able to move relative to each other, resulting in a conformational change that is two-fold symmetric (Fig. 3f). It may be that the change in the angle of the cleft between the N1–N2 lobes represents a conserved response of RCK domains to ligand binding that introduces a strain on the gating ring, and that differences in the architecture and strength of the intersubunit or assembly interfaces determines how this strain is relieved, resulting in an overall conformational change to the ring that is suitable for the particular associated channel.

Mechanistic implications for channel regulation by TrkA

Comparison of the two TrkA conformations suggests a possible mechanism for opening of the channel by nucleotide-induced conformational changes on the gating ring. In the closed state bound to ADP, probably corresponding to the structure of the TrkH–TrkA complex, the ion permeation pathway is blocked by the intramembrane loop connected to helix D3M2b. Although the ‘twisted’ conformation of TrkA was obtained in the absence of the channel, possible effects of this conformational change on the channel can be probed by mapping the locations of residues forming the interface in the TrkH–TrkA structure onto the TrkA-only structure (Fig. 4a). The surface residues interacting with the channel are located primarily in two patches on the diagonally positioned N2 domains facing the channel, each of which interacts with one of the channel D3M2b helices. These two interface regions are farther apart by ~24 Å in the isolated TrkA structure (Fig. 4a and Supplementary Video 5). If ATP binding causes the gating ring to adopt a twisted conformation similar to the free TrkA structure, separation of the two interfaces on the tetramer could pull the helix D3M2b away from the centre of the

ion permeation pathway, leading to movement of the loop out of the pore and opening of the channel (Fig. 4b). As a first test of whether the intramembrane loop is in fact required for the complex’s response to ATP, currents were measured from the complex of the wild-type TrkA and a TrkH mutant with residues 348–354 of the intramembrane loop deleted. The loop deletion mutant had a higher basal open probability than the wild-type channel, and was largely insensitive to perfusion of ATP (Fig. 4c). Although this result is consistent with the hypothesized role of the intramembrane loop in gating of TrkH by TrkA, further structural and functional studies are necessary to reveal the molecular mechanism governing regulation of channel activity.

Discussion

This study presents, to our knowledge, the first demonstration that TrkH, and by extrapolation the TrkG and KtrB family of proteins, are ion channels. Additionally, we demonstrate that ion flux through the TrkH–TrkA complex is upregulated by ATP and downregulated by ADP. It is likely that in a cell the ratio of ATP to ADP determines the activity of TrkH, thus providing a way for regulating ion conduction in response to the metabolic activity of an organism. However, it is puzzling that a protein that promotes growth of bacteria in low external K^+ concentrations only has a modest selectivity between K^+ and Na^+ .

One curious feature of the Na^+ and Li^+ currents is that the amplitude of level 2 was not twice that of level 1. Similar behaviour was observed for the I220C channels modified by MTSET (Supplementary Figs 1, 2 and Supplementary Table 1). One possible interpretation of this result is that under these conditions the conductance through an individual channel is lower when its partner is closed than when both are open. Because the D3M2b helix is connected to the intramembrane loop and also forms significant contacts with the neighbouring protomer, it represents a possible avenue for coupling of gating in one

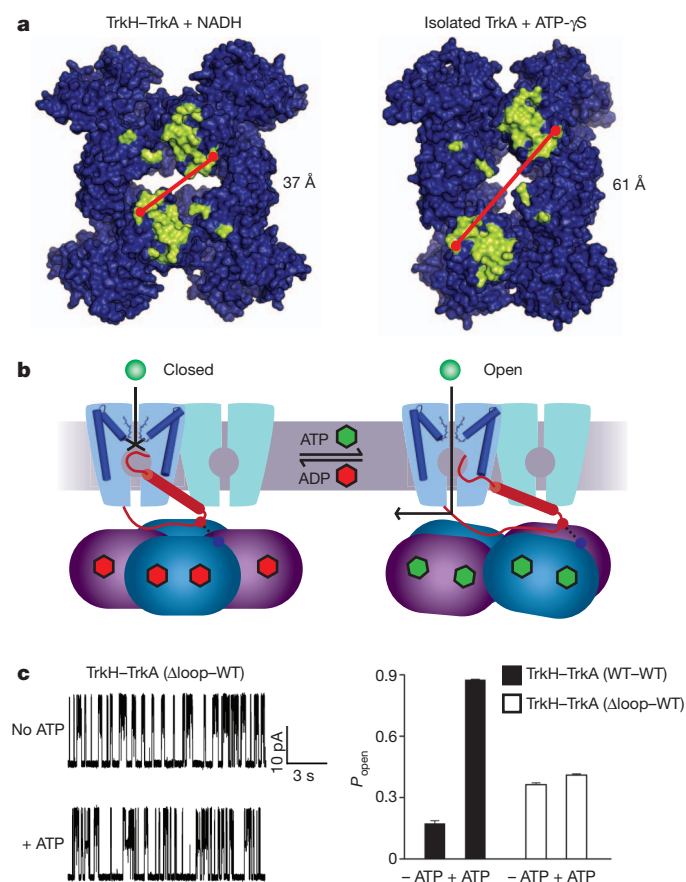


Figure 4 | Proposed mechanism of regulation of TrkH gating by TrkA. **a**, Surface representation of the TrkA tetramer from the TrkA-only (top) and TrkH-TrkA structures, viewed from the membrane-facing side. Residues forming the channel-gating ring interface, defined as residues with at least one atom within 4 Å of TrkH in the complex structure, are marked in yellow. **b**, Diagram illustrating a possible gating mechanism, with the closed channel shown on the left, and the open channel on the right. The D3M2b helix is represented as a red cylinder. For clarity, gating of only one TrkH protomer is shown. **c**, Left panel: single-channel currents through the TrkH-TrkA (Δloop-WT) complex before and after addition of ATP. The holding potential is +50 mV. Right panel: the P_{open} of TrkH-TrkA (WT-WT) and TrkH-TrkA (Δloop-WT) before and after addition of 5 mM ATP. The error bars are s.e.m. from three independent patches.

channel to conductance in its partner. However, why this behaviour occurs more prominently when Na^+ or Li^+ occupies the selectivity filter remains unknown.

In addition, we present two different structures of the TrkA gating ring, solved in complex with TrkH or in isolation. From these structures we have proposed a model for the gating of TrkH by TrkA that assigns the conformation in the complex to the closed state of TrkA, and the conformation in isolation to the open state. However, although the 'open' structure was obtained bound to an analogue of the channel opener ATP, it should be noted that crystal packing and the absence of the channel probably have a significant effect on the gating ring's energy landscape, and these factors rather than the identity of the bound ligand may be responsible for the conformation observed in the crystal structure of the isolated TrkA tetramer. In fact, a recently reported structure of TrkA from *Vibrio vulnificus* (Protein Data Bank accession 4G65) assumes a conformation highly similar to the twisted conformation reported here despite containing no bound nucleotide. Nonetheless, although it may not correspond exactly to the structure in the open complex, the isolated TrkA structure reveals the range of flexibility accessible to the gating ring. Confirmation for the proposed model will require further structural and functional studies of the intact complex.

METHODS SUMMARY

Vibrio parahaemolyticus TrkH (GenBank accession EDM57342.1) and TrkA (EDM59790.1) were overexpressed in BL21(DE3) cells and purified by affinity chromatography, followed by cleavage of the tag and a second purification step on a size-exclusion column. For purification of the TrkH-TrkA complex, TrkA was expressed with a C-terminal polyhistidine affinity tag; for the purification of isolated TrkA, TrkA was expressed with an N-terminal GST domain as an affinity tag. The TrkH-TrkA complex was solubilized in 40 mM *n*-decyl-β-D-maltoside, which was decreased to 3.5 mM during the size-exclusion chromatography step. TrkH-TrkA crystals were grown by the microbatch method and derivatized with $\text{Ta}_6\text{Br}_{12}$ before freezing. TrkA crystals were grown in the presence of 10 mM ATP-γS by the sitting-drop vapour diffusion method. The structures of TrkA and the complex were solved using molecular replacement or combined molecular replacement/single-wavelength anomalous dispersion, respectively, with the PHENIX software suite²⁸.

Patch-clamp recording of the TrkH-TrkA complex was performed on *E. coli* giant spheroblasts as described previously^{29–31}. Single-channel currents were recorded on inside-out patches, with both the bath solution and the pipette solution containing 200 mM KCl, 0.5 M sucrose, and 10 mM HEPES at pH 7.4. The data were analysed with pClampfit9.

Full Methods and any associated references are available in the online version of the paper.

Received 4 October 2012; accepted 5 March 2013.

- Corratgé-Faillie, C. *et al.* Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. *Cell. Mol. Life Sci.* **67**, 2511–2532 (2010).
- Cao, Y. *et al.* Crystal structure of a potassium ion transporter, TrkH. *Nature* **471**, 336–340 (2011).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Hänelt, I. *et al.* Membrane region M2C2 in subunit KtrB of the K^+ uptake system KtrAB from *Vibrio alginolyticus* forms a flexible gate controlling K^+ flux: an electron paramagnetic resonance study. *J. Biol. Chem.* **285**, 28210–28219 (2010).
- Hänelt, I. *et al.* Gain of function mutations in membrane region M2C2 of KtrB open a gate controlling K^+ transport by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.* **285**, 10318–10327 (2010).
- Bossemeyer, D. *et al.* K^+ -transport protein TrkA of *Escherichia coli* is a peripheral membrane protein that requires other *trk* gene products for attachment to the cytoplasmic membrane. *J. Biol. Chem.* **264**, 16403–16410 (1989).
- Nakamura, T., Yuda, R., Unemoto, T. & Bakker, E. P. KtrAB, a new type of bacterial K^+ -uptake system from *Vibrio alginolyticus*. *J. Bacteriol.* **180**, 3491–3494 (1998).
- Schlosser, A., Hamann, A., Bossemeyer, D., Schneider, E. & Bakker, E. P. NAD^+ binding to the *Escherichia coli* K^+ -uptake protein TrkA and sequence similarity between TrkA and domains of a family of dehydrogenases suggest a role for NAD^+ in bacterial transport. *Mol. Microbiol.* **9**, 533–543 (1993).
- Kong, C. *et al.* Distinct gating mechanisms revealed by the structures of a multi-ligand gated K^+ channel. *eLife* **1**, e00184 (2012).
- Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. Structure of the RCK domain from the *E. coli* K^+ channel and demonstration of its presence in the human BK channel. *Neuron* **29**, 593–601 (2001).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Roosild, T. P., Miller, S., Booth, I. R. & Choe, S. A mechanism of regulating transmembrane potassium flux through a ligand-mediated conformational switch. *Cell* **109**, 781–791 (2002).
- Roosild, T. P. *et al.* KTN (RCK) domains regulate K^+ channels and transporters by controlling the dimer-hinge conformation. *Structure* **17**, 893–903 (2009).
- Albright, R. A., Ibar, J. L., Kim, C. U., Gruner, S. M. & Morais-Cabral, J. H. The RCK domain of the KtrAB K^+ transporter: multiple conformations of an octameric ring. *Cell* **126**, 1147–1159 (2006).
- Wu, Y., Yang, Y., Ye, S. & Jiang, Y. Structure of the gating ring from the human large-conductance Ca^{2+} -gated K^+ channel. *Nature* **466**, 393–397 (2010).
- Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y. & MacKinnon, R. Structure of the human BK channel Ca^{2+} -activation apparatus at 3.0 Å resolution. *Science* **329**, 182–186 (2010).
- Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossman, M. G. D-glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance. *Proc. Natl Acad. Sci. USA* **70**, 3052–3054 (1973).
- Roosild, T. P. *et al.* Mechanism of ligand-gated potassium efflux in bacterial pathogens. *Proc. Natl Acad. Sci. USA* **107**, 19784–19789 (2010).
- Ye, S., Li, Y., Chen, L. & Jiang, Y. Crystal structures of a ligand-free MthK gating ring: insights into the ligand gating mechanism of K^+ channels. *Cell* **126**, 1161–1173 (2006).
- Tholema, N. *et al.* All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K^+ uptake by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.* **280**, 41146–41154 (2005).
- Matsuda, N. *et al.* Na^+ -dependent K^+ uptake Ktr system from the cyanobacterium *Synechocystis* sp. PCC 6803 and its role in the early phases of cell adaptation to hyperosmotic shock. *J. Biol. Chem.* **279**, 54952–54962 (2004).

22. Bakker, E. P. & Mangerich, W. E. The effects of weak acids on potassium uptake by *Escherichia coli* K-12 inhibition by low cytoplasmic pH. *Biochim. Biophys. Acta* **730**, 379–386 (1983).
23. Yuan, P., Leonetti, M. D., Hsiung, Y. & MacKinnon, R. Open structure of the Ca^{2+} gating ring in the high-conductance Ca^{2+} -activated K^{+} channel. *Nature* **481**, 94–97 (2012).
24. Akabas, M. H., Stauffer, D. A., Xu, M. & Karlin, A. Acetylcholine receptor channel structure probed in cysteine-substitution mutants. *Science* **258**, 307–310 (1992).
25. Rhoads, D. B. & Epstein, W. Energy coupling to net K^{+} transport in *Escherichia coli* K-12. *J. Biol. Chem.* **252**, 1394–1401 (1977).
26. Stewart, L. M., Bakker, E. P. & Booth, I. R. Energy coupling to K^{+} uptake via the Trk system in *Escherichia coli*: the role of ATP. *J. Gen. Microbiol.* **131**, 77–85 (1985).
27. Wierenga, R. K., Terpstra, P. & Hol, W. G. Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187**, 101–107 (1986).
28. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
29. Long, W. S., Slayman, C. L. & Low, K. B. Production of giant cells of *Escherichia coli*. *J. Bacteriol.* **133**, 995–1007 (1978).
30. Felle, H., Porter, J. S., Slayman, C. L. & Kaback, H. R. Quantitative measurements of membrane potential in *Escherichia coli*. *Biochemistry* **19**, 3585–3590 (1980).
31. Kuo, M. M., Saimi, Y., Kung, C. & Choe, S. Patch clamp and phenotypic analyses of a prokaryotic cyclic nucleotide-gated K^{+} channel using *Escherichia coli* as a host. *J. Biol. Chem.* **282**, 24294–24301 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements Diffraction data for this study were measured at beamlines X4A, X4C, X25 and X29 of the National Synchrotron Light Source, ID-17 and ID-24 at the Advanced Photon Source, 5.0.2 and 8.2.2 at the Advanced Light Source, and A1 and F1 at the Cornell High Energy Synchrotron Source. M.Z. thanks E. Bakker, I. Booth, R. Krämer and C. Slayman for discussions and insights, and K. R. Rajashankar for help with X-ray crystallography. This work was supported by the US National Institutes of Health (DK088057, GM098878 and HL086392 to M.Z., and a subcontract (sub 5808 to M.Z.) from the PSI: Biology grant U54GM095315 to W. Hendrickson), the American Heart Association (12EIA8850017 to M.Z. and 0826067D to Y.P.), and the Irma T. Hirschl Trust Award to M.Z. M.Z. was a Pew Scholar in Biomedical Sciences.

Author Contributions Y.C., Y.P., H.H., E.J.L., X.J. and M.Z. designed the experiments. Y.C. and H.H. expressed, purified and crystallized the proteins, and collected and processed the diffraction data. H.H., X.J. and E.J.L. solved and refined the structures. Y.P. prepared spheroplasts and performed patch-clamp recordings. B.K. provided reagents and the VpTrkH clone. All authors participated in data analysis. Y.P., E.J.L. and M.Z. wrote the manuscript with inputs from all authors.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession numbers 4J9U and 4J9V. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Z. (mzhou@bcm.edu).

METHODS

Cloning and purification of the TrkH–TrkA complex. The *trkH* gene was amplified from *Vibrio parahaemolyticus* genomic DNA and was cloned into the pET31b plasmid (Novagen) with an ampicillin-resistance gene. The *Vibrio parahaemolyticus trkA* gene was cloned into a modified pET plasmid (Novagen) with a C-terminal polyhistidine tag and a TEV protease recognition site with a kanamycin-resistance gene. For large-scale purification of the VpTrkH–TrkA complex, BL21(DE3) cells were co-transformed with a 1:1.5 ratio of plasmids carrying VpTrkH and VpTrkA, respectively. The transformants were grown in Luria broth supplemented with 100 mg l^{−1} kanamycin and 100 mg l^{−1} ampicillin at 37 °C and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) after the optical density at 600 nm reached 1.0. The cell membranes were solubilized with 40 mM *n*-decyl-β-D-maltoside (Anatrace) and the TrkH–TrkA complex was purified with TALON Metal Affinity Resin (Clontech Inc.). After removal of the His tag with TEV protease, the TrkH–TrkA was subjected to size-exclusion chromatography with a Superdex 200 10/300 GL column (GE Health Sciences) pre-equilibrated in a buffer of 150 mM KCl, 20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol and 3.5 mM *n*-decyl-β-D-maltoside. For crystallization, the complex was concentrated to 10 mg ml^{−1} as approximated by ultraviolet absorbance and NADH was added to a final concentration of 4 mM.

Cloning and purification of isolated TrkA. For expression of VpTrkA alone, the VpTrkA gene was cloned into a modified pGEX-5x-1 GST expression vector (GE lifescience) with a TEV protease recognition site replacing the original Factor Xa site. The plasmid was transformed into BL21(DE3) cells, and the transformants were grown in Luria broth supplemented with 100 mg l^{−1} ampicillin at 37 °C. Once the OD_{600 nm} reached 1.0, expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20 °C for 15 h. GST–VpTrkA was purified with glutathione Sepharose 4B resin (Amersham Bioscience). After removal of the GST with TEV protease, VpTrkA was subjected to size-exclusion chromatography with a Superdex 200 10/300 GL column (GE Health Sciences) pre-equilibrated in a buffer of 150 mM KCl, 20 mM HEPES, pH 7.5 and 5 mM β-mercaptoethanol. The protein was concentrated to 8 mg ml^{−1} as approximated by BCA method.

Crystallization and structure solution. TrkH–TrkA crystals were grown by microbatch crystallization under mineral oil where 1.5 μl of the protein solution was mixed with an equal volume of crystallization solution containing 37.5% PEG400, 400 mM ammonium sulphate and 100 mM HEPES, pH 6.75. Before collecting crystals, tantalum bromide clusters (Jena Bioscience) were applied by dissolving approximately 25 μg μl^{−1} powder directly into the crystal growth solution and incubating for 10–12 h. Before flash-freezing in liquid nitrogen, the crystals were cryo-protected by soaking for 2–5 s in a solution obtained by sitting-drop vapour diffusion with equal volumes of protein solution and well solution containing 35% PEG400, 400 mM ammonium sulphate, 100 mM HEPES, pH 6.75 and 25 μg μl^{−1} Ta₆Br₁₂.

Isolated VpTrkA was co-crystallized with ATP-γS by vapour diffusion in sitting drops, where 1.5 μl of the protein solution supplemented with 10 mM ATP-γS was mixed with an equal volume of crystallization solution. Crystals were obtained in two conditions: 10% polyacrylic acid 5100, 100 mM Tris, pH 8.0, 20 mM MgCl₂ and 5% glycerol; or 30% PEG400, 100 mM glycine, pH 9.0 and 200 mM MgSO₄. Before flash-freezing in liquid nitrogen, the crystals were cryo-protected by increasing the concentration of glycerol (25% v/v) and PEG400 (35% v/v) in the mother liquor.

X-ray diffraction data were collected at the beamlines X29 and X4A and C at the National Synchrotron Light Source, ID-24 and ID-17 at the Advanced Photon Source, 5.0.2 and 8.2.2 at the Advanced Light Source, and A1 and F1 at CHESS. The structure of the TrkH–TrkA complex was solved by MRSA using the Phenix software suite²⁸, with the TrkH structure (3PJZ) as a search model and experimental phases from the 18 Ta₆Br₁₂ clusters in the asymmetric unit. Iterative rounds of manual and automatic structure refinement were carried out with Coot³² and phenix.refine, respectively, and protein geometry was monitored using Molprobity³³. The final model contains 18 tantalum bromide clusters, 4 NADH molecules, 4 TrkA molecules and 4 TrkH molecules. Residues 63–65 and 159–177, located on cytoplasmic loops, are unresolved in the TrkH molecules; residues 162–163 and 178–181 are unresolved in TrkA. The isolated TrkA structure was solved using molecular replacement with the TrkA structure from the complex, and refined by a similar method to the structure of the complex. The final model of the isolated TrkA structure contained four molecules of ATP-γS and two TrkA molecules.

Electrophysiology. The VpTrkH- and VpTrkA-containing vectors used in expressing the TrkH–TrkA complex were used. All mutations were made with the QuikChange kit (Agilent Technologies) and verified by sequencing through the entire coding regions. Giant spheroplasts were prepared from *E. coli* following a protocol described previously^{29–31}, with minor modifications. In brief, a single colony co-transformed with TrkH and TrkA was incubated in 5 ml of modified LB medium (0.5% NaCl, 1% tryptone and 0.5% yeast extract) by shaking at 250 r.p.m., 37 °C. When the OD₆₀₀ reached 0.5, the culture was diluted by 10-fold into 5 ml fresh modified LB with the addition of 60 μg ml^{−1} cephalixin. The elongated filaments were then collected by centrifuging 2 ml of the culture at 3,000g for 1 min. The supernatant was removed carefully and 0.5 ml 0.8 M sucrose was added to re-suspend the pellet. The following solutions were added sequentially: 30 μl of 1 M Tris-HCl (pH 8.0), 24 μl of 0.5 mg ml^{−1} lysosome, 6 μl of 5 mg ml^{−1} DNase I, and 6 μl of 0.125 M EDTA-NaOH (pH 8.0). The mixture was incubated for 7.5 min at room temperature and 100 μl of the stop solution (0.7 M sucrose, 20 mM MgCl₂, and 10 mM Tris-HCl, pH 8.0) was added to stop the digestion. The spheroplasts were aliquoted and stored at −20 °C. The frozen spheroplasts were thawed on ice and used in electrophysiology.

Patch-clamp currents were recorded on inside-out patches pulled from the spheroplasts placed on clean glass coverslips. Electrodes were drawn from patch glass (TW100F-4, World Precision Instruments) and polished (MP-803, Narishige Co.) to a resistance of 7–10 MΩ. Both bath solution and the pipette solution contained 200 mM KCl, 0.5 M sucrose and 10 mM HEPES at pH 7.4. All solutions of perfused nucleotides were prepared from sodium salts. No pressure was applied to the membrane after the GΩ seal formation was achieved, which excluded the endogenous mechanosensitive channel activities³¹. The analogue signals were amplified by an Axon-200B patch-clamp amplifier (Molecular Devices Inc.), and filtered at 1 kHz using the built-in Bessel filter, digitized at 100 μs by Digidata 1322a (Molecular Devices Inc.), and recorded to a computer hard disk. The data were analysed with pClampfit9 software.

32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
33. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

The structure of the KtrAB potassium transporter

Ricardo S. Vieira-Pires¹, Andras Szollosi¹ & João H. Morais-Cabral¹

In bacteria, archaea, fungi and plants the Trk, Ktr and HKT ion transporters are key components of osmotic regulation, pH homeostasis and resistance to drought and high salinity. These ion transporters are functionally diverse: they can function as Na⁺ or K⁺ channels and possibly as cation/K⁺ symporters. They are closely related to potassium channels both at the level of the membrane protein and at the level of the cytosolic regulatory domains. Here we describe the crystal structure of a Ktr K⁺ transporter, the KtrAB complex from *Bacillus subtilis*. The structure shows the dimeric membrane protein KtrB assembled with a cytosolic octameric KtrA ring bound to ATP, an activating ligand. A comparison between the structure of KtrAB–ATP and the structures of the isolated full-length KtrA protein with ATP or ADP reveals a ligand-dependent conformational change in the octameric ring, raising new ideas about the mechanism of activation in these transporters.

The Trk/Ktr/HKT superfamily of ion transporters includes the Trk, Ktr and HKT transporter families¹. These transporters are closely related to the superfamily of tetrameric cation channels^{2–4}, which includes potassium, sodium and calcium channels. The membrane proteins in both superfamilies share the same architecture^{2,5–9}: four subunits or repeats, each with the TM–P loop–TM (where TM indicates transmembrane helix) structural motif first seen in the KcsA potassium channel structure¹⁰, assemble to form an ion pore with a ‘selectivity filter’ along its central axis. In addition, the cytosolic regulatory proteins of the Trk and Ktr transporters and of the MthK^{11,12} and large-conductance Ca²⁺-gated (BK) K⁺ channels^{13–15} are RCK (regulate conductance of K⁺) domains.

Ktr ion transporters are crucial K⁺ transport systems in some bacteria^{16,17}, with a role in resistance to osmotic stress and high salinity by mediating the early uptake of K⁺ (refs 16, 18). The Ktr ion transporters are composed of two essential components^{3,16}: a membrane protein (KtrB or KtrD) and a cytosolic regulatory protein (KtrA or KtrC). Studies with cells show that Ktr proteins^{8,19} transport K⁺ but are also permeable to Na⁺, and that K⁺ transport is ATP²⁰ and Na⁺ dependent^{5,8,18}. Furthermore, it has been shown that truncated KtrA forms octameric rings^{21,22} and that KtrB assembles as homodimers^{6,21}. The structures described here clarify the molecular basis of some of these properties as well as the structural relationship between these transporters and ion channels.

KtrAB architecture

We have determined the structure of a KtrAB potassium transporter with diffraction data to 3.5 Å (Fig. 1 and Supplementary Table 1). Two diffraction data sets to better than 4 Å were collected out of many thousands of crystals tested. The structure was determined from an anisotropic diffraction data set which extends to 3.3 Å in the best direction; four- and eight-fold averaging together with cross-crystal averaging led to the calculation of high-quality electron-density maps (Supplementary Fig. 1). The crystallographic asymmetric unit contains a KtrA octameric ring associated with two identical KtrB dimers, one on each face of the ring (Supplementary Fig. 2). This arrangement resulted from the *in vitro* assembly of independently expressed and purified protein components and the existence of two identical KtrB binding surfaces in KtrA.

In the bacterial membrane the KtrAB complex is composed of a homodimeric KtrB membrane protein and a cytoplasmic octameric KtrA ring (Fig. 1a). The KtrA octameric ring (Fig. 1a, b and Supplementary Fig. 3) is a four-fold symmetric assembly of four KtrA dimers with the typical RCK domain arrangement^{12,13,15}, the amino (N) lobe forming the ring and the smaller carboxy (C) lobe at the periphery. The ligand-binding site in the N lobe is occupied by an ATP molecule (Fig. 1a and Supplementary Figs 1b and 3b). The inner circumference of the ring delimits a ~30-Å-wide hole which, together with gaps (approximate dimensions: 17 Å by 9 Å and 18 Å by 11 Å) between the surfaces of the KtrB homodimer and the KtrA ring, give access from the cytosol to the cytoplasmic pore of KtrB (Fig. 1b and Supplementary Fig. 3a).

The interactions between KtrB and the KtrA ring are an important difference relative to K⁺ channels^{11–15}. In channels the pore subunit and one or two RCK domains are encoded in the same polypeptide and the RCK octameric ring hangs from the cytoplasmic face of the membrane protein through a polypeptide tether. In the KtrAB complex, the KtrB homodimer sits directly on a face of the KtrA ring (Fig. 1c), close to a ring diagonal, and cytoplasmic loops establish two different types of contacts with pairs of RCK domains in the ring: the lateral contact and the tip contact (Fig. 1c, d). The lateral-contact regions involve a mid-section of the KtrB carboxy terminus, which runs along the homodimer interacting with the neighbouring KtrB subunit and the KtrA ring (Fig. 1c, d and Supplementary Fig. 4a). The tip-contact regions occur at the apexes of the cytoplasmic face of the KtrB homodimer (Fig. 1c, d and Supplementary Fig. 4b) where the KtrB cytoplasmic loops connecting repeats D1 and D2 interact with two opposing KtrA subunits. The functional relevance of these interactions and of our *in vitro* assembled KtrAB complex is demonstrated in the Supplementary Discussion.

KtrB structure

The KtrB subunit structure has four repeats (D1 to D4) wrapped around a central axis and cradling the selectivity filter (Figs 1c and 2a). Each repeat (Supplementary Figs 5a, b) includes the M1 helix spanning the bilayer, followed by the P-loop region (that includes the pore (P) helix and the selectivity-filter sequence) and a second shorter (M2a) helix, which ends close to the centre of the membrane. A C-terminal (M2b)

¹IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, Porto 4150-180, Portugal.

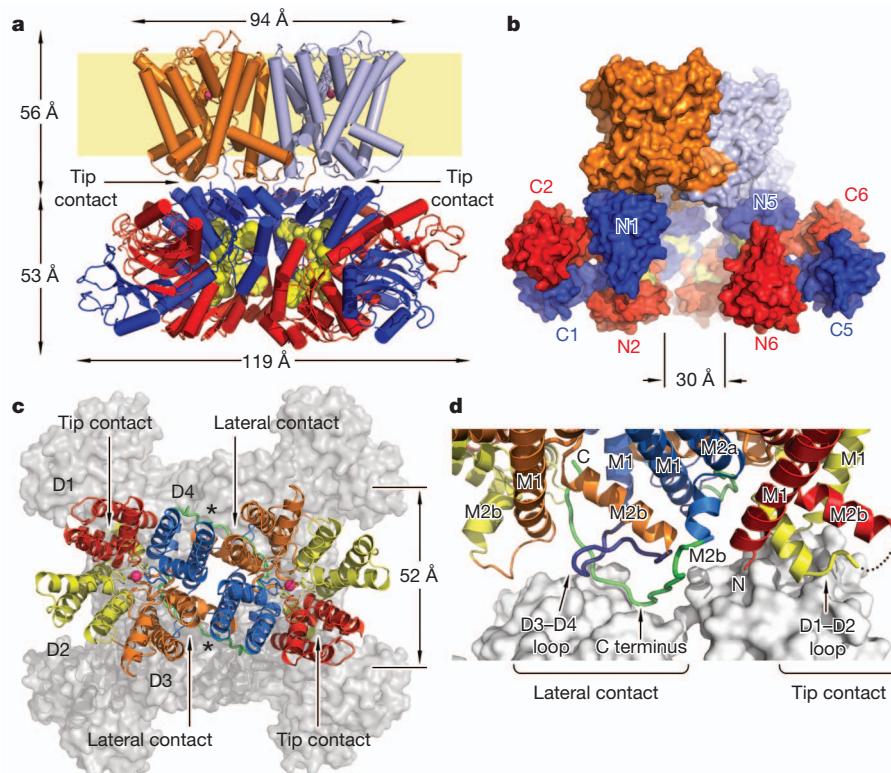


Figure 1 | Structure of the KtrAB K^+ transporter. **a**, Cartoon of the KtrAB structure with the extracellular side at the top. KtrB subunits are in orange and light blue whereas KtrA subunits are in dark blue and red; ATP is shown as yellow spheres and K^+ as magenta spheres; cell membrane is shown as a pale yellow rectangle. **b**, Surface representation of KtrAB rotated relative to panel **a** along the vertical axis. One KtrA dimer has been removed for viewing the hole at centre of the octameric ring. N and C lobes of KtrA subunits are indicated.

helix completes each repeat, running at a shallow angle towards the membrane cytoplasmic face. The selectivity filter is funnel-shaped with its wider mouth (11–13 Å across) towards the extracellular face and the single ion-binding site forming the narrow end (Fig. 2a and Supplementary Fig. 5b). Immediately below the KtrB selectivity filter there is an intramembrane loop (Fig. 2a–c), a residue stretch that connects helices M2a to M2b in repeat D3 and is composed almost exclusively of glycines, alanines and serines (Supplementary Fig. 5c). The intramembrane loop together with a highly conserved arginine (R417) form a structure that blocks access between the selectivity filter and the cytoplasmic pore; mutations or truncations in this region increase ion flux and this region has been proposed to be a transporter gate^{9,19}. An open pathway runs from the R417/intramembrane loop to the cytoplasmic face of the protein (Fig. 2a). The pathway is lined by a mix of polar and apolar residues and is ~4 Å wide at its narrowest point. Overall, this architecture resembles closely the architecture of potassium channel pores (Supplementary Fig. 6).

The monomer and homodimer structures of KtrB and of the related TrkH⁹ (a Trk membrane protein) are very similar (Supplementary Fig. 7). A major difference between KtrB and TrkH is found in the C terminus. In KtrB the long C terminus runs from one subunit, forms a lateral contact and snakes into the cytoplasmic pore of the neighbouring KtrB, finishing just below the intramembrane loop (Figs 1c, d and 2a, b). In contrast, the TrkH C terminus is much shorter and does not establish any contacts (Supplementary Figs 5c and 7a, b). The last seven residues of the KtrB C terminus (Supplementary Fig. 1c) line the wall of the cytoplasmic pore of the neighbouring subunit (Fig. 2a, b). The highly conserved C-terminal glycine (G445) is positioned with its carboxylate interacting with the side chain of a lysine, which is part of the intramembrane loop and is

c, Extracellular view of KtrB repeats (D1, red; D2, yellow; D3, orange; D4, blue); KtrA is shown as grey surface. Asterisks indicate green KtrB C termini running along the KtrB dimer. **d**, Detail of contact regions. Colouring as in panel **c** except for the D3–D4 loop, shown in purple. The disordered stretch in the D1–D2 loop is shown as a dotted line. KtrB helices and N and C termini are labelled.

conserved in both the Ktr (K315 in KtrB) and Trk (K357 in TrkH) transporters (Fig. 2c). The C-terminal glycine does not exist in Trk proteins, but in the TrkH structure the side-chain carboxylate of a conserved glutamate (E470) occupies the same volume as the glycine carboxylate and also interacts with the lysine.

To assess the importance of positioning a carboxylate close to the intramembrane loop we generated truncations of the C-terminal residues and evaluated their impact in a growth complementation assay^{3,6,8} of the TK2420 *Escherichia coli* strain. This strain has its major K^+ transport systems disabled and only grows in K^+ concentration ≥ 30 mM; however, when expressing wild-type KtrAB it can grow in K^+ concentrations of ~0.3 mM. The position of the carboxylate does not seem to be essential as sequential truncation of up to three C-terminal residues has no effect or causes only a small shift in rescue (Fig. 2d). Furthermore, the stability of the KtrAB complex seems to be undisturbed (Fig. 2e). However, removal of four residues results in a non-functional transporter, equivalent to control plasmid (Fig. 2d). Consistently, KtrB without the last four residues, although it still forms a homodimer, does not assemble with KtrA (Fig. 2e). All mutations display similar expression levels (Supplementary Fig. 4f). Presumably, the fourth residue, a conserved apolar residue which in the structure seems to function as a C terminus anchor, has a role in the positioning of the C terminus and therefore on the stability of the C-terminus-mediated lateral contact. Notably, we had previously shown that truncation of the last 10 or 15 residues of the C terminus abolishes strain complementation and destabilizes the KtrB dimer⁶. Although we cannot rule out more subtle functional roles for the C terminus, such as cooperativity between KtrB subunits, our results demonstrate that the C terminus has a role both in the stability of the KtrB homodimer and of the KtrAB complex.

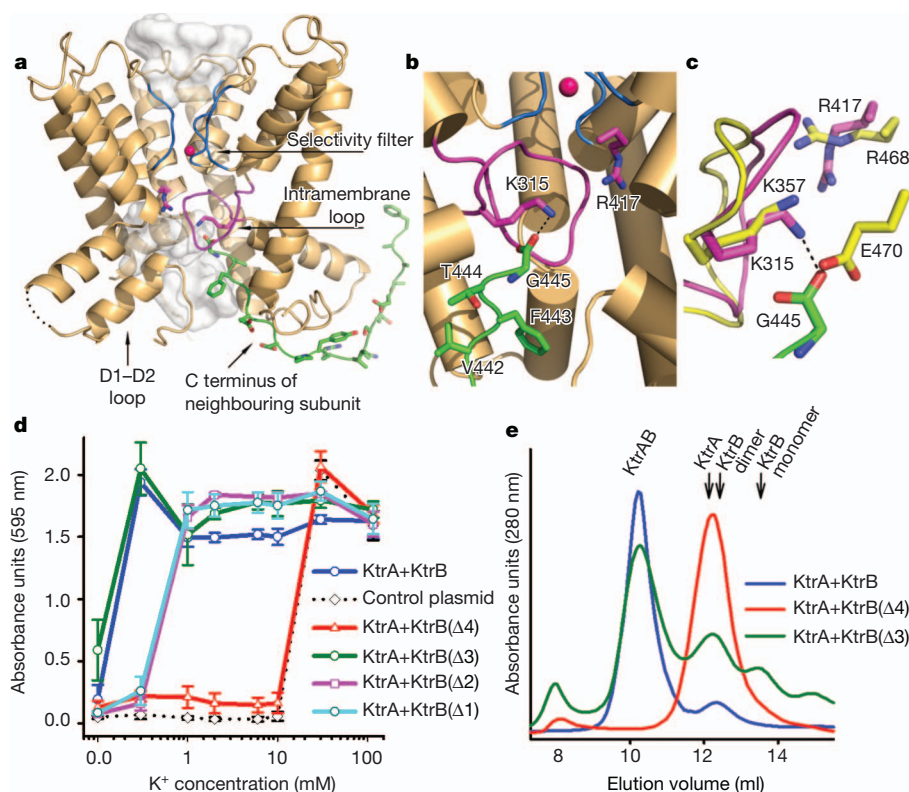


Figure 2 | KtrB subunit. **a**, KtrB subunit without repeat D2. K^+ ion is indicated as a red sphere; the intramembrane loop, K315 (at the centre) and R417 (on the left) are shown in magenta. The ion pathway, calculated with Hollow⁴¹ with a 1.4 Å probe radius, is shown as a white surface. **b**, Close-up view of ion pathway, rotated 180° relative to panel **a**. C-terminal residues from the neighbouring KtrB subunit are shown in green stick representation. **c**, Intramembrane loops in KtrB (magenta and green) and TrkH (yellow), after

superposition of pore regions. **d**, Optical density of *E. coli* TK2420 (expressing different constructs) cultures versus K^+ concentration. $\Delta 1$, $\Delta 2$, $\Delta 3$ and $\Delta 4$ indicate KtrB C-terminal sequential truncations. Mean \pm s.d. from three independent experiments is shown. **e**, Size-exclusion chromatography elution profiles of mixtures of wild-type KtrA and wild-type or mutant KtrB, at same concentrations and ratios. Elution volumes for KtrAB, KtrA (octamer) and KtrB (monomer and dimer) are indicated.

Regulation of KtrAB activity

We first determined that KtrA has a strong binding preference for ATP and ADP over other small molecules (Fig. 3a) resembling other KtrA proteins²⁰. The functional effect of these two ligands was then evaluated in a $^{86}\text{Rb}^+$ uptake assay^{6,9,23} by reconstituting KtrAB with ATP or ADP (Supplementary Fig. 8a) in the presence of an excess of KtrA to minimize the formation of the KtrB–KtrA–KtrB complex. KtrB alone was also reconstituted; it allows K^+ into cells but less efficiently than KtrAB⁸.

All protein forms mediate flux in this assay (Fig. 3b). Importantly, flux is markedly higher with ATP, supporting previous cell-based results²⁰ which showed that ATP is an activator of KtrAB. Noticeably, the $^{86}\text{Rb}^+$ flux difference between KtrAB–ATP and KtrB (~ 2 -fold difference at 30 min) is smaller than expected from K^+ uptake measurement in cells⁸ (~ 10 -fold difference in V_{max}), which may indicate that, in addition to ATP, other activating factors are necessary for full transporter activity.

To understand how the two ligands affect KtrA we determined the crystal structures of this protein bound to ATP and ADP at ~ 3.2 Å and ~ 2.9 Å, respectively (Fig. 4a, b and Supplementary Table 1). The ADP- and ATP-bound octameric rings adopt different conformations: the KtrA–ATP ring adopts a four-fold symmetric square conformation and the KtrA–ADP ring adopts a two-fold symmetric diamond conformation. A conformational difference is also observed in solution, through controlled proteolysis (Supplementary Fig. 8b). Both structures present clear electron density for the ligand's adenosine and phosphate groups (Supplementary Fig. 1d, e). Although the conformation of the adenosine group and its interactions with the protein are very similar in the two structures, the phosphate groups

are rearranged (Fig. 4c); in particular, the γ -phosphate in the ATP structure occupies the position taken by the β -phosphate in the ADP structure. This is accomplished by a shift in G79 and a repositioning of the α - and β -phosphates relative to the ADP structure. The phosphate–protein interactions are also different: in the ADP structure the β -phosphate interacts with residues from its own binding site, in particular with the side chains R16 and K103; in contrast, in the ATP structure the γ -phosphate interacts with R16 from its binding site, whereas the β -phosphate interacts with R16 from the neighbouring KtrA subunit binding site (Fig. 4c). Presumably owing to the intradimer interactions mediated by ATP, the two binding sites within a KtrA dimer are brought together through a $\sim 16^\circ$ change in the angle between subunits (Fig. 4d); the distances separating the conserved D36 in the two binding pockets are ~ 35 Å in the ADP structure and ~ 30 Å in the ATP structure. This dimer rearrangement is accompanied by rigid-body rotations across two of the four dimer-to-dimer interfaces in the ring and is reflected in the different ring conformations. Most notably the distance separating F71 (a KtrA residue that participates in the KtrAB tip contact region) in opposite ring subunits is 68.7 Å in the KtrA–ATP ring, whereas in KtrA–ADP the equivalent distances are 68.3 and 81.8 Å (Fig. 4a, b). Similarly, the distance separating L66 (a KtrA residue that participates in the KtrAB lateral contact region) in KtrA–ATP is 49.8 Å, whereas the equivalent distances in KtrA–ADP are 49.8 and 66.1 Å. Therefore, the conformational change from the KtrA–ADP ring to the ATP ring involves the asymmetric contraction of the ring, with one diagonal direction changing by 13–16 Å, whereas the other diagonal is unaffected.

Notably, the conformation of the octameric ring and the details of ligand binding are basically identical between the isolated KtrA–ATP

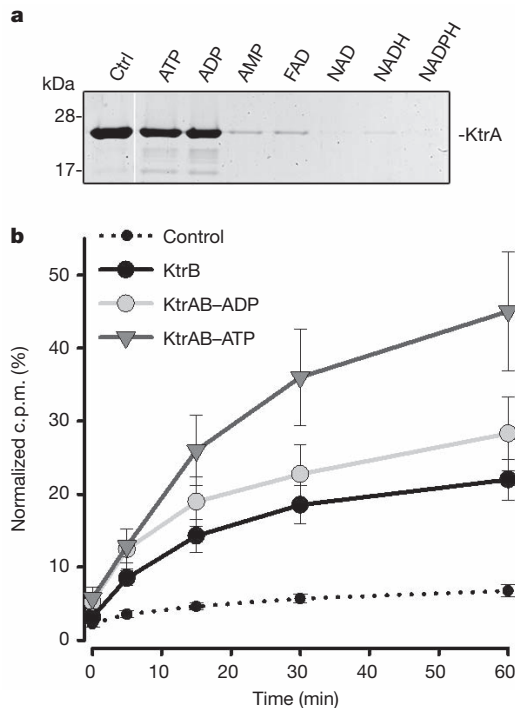


Figure 3 | Ligand regulation of KtrAB transporter. **a**, SDS-PAGE of KtrA protein eluted from ADP-agarose with 1 mM of different adenosine-containing ligands. Purified KtrA is indicated as Ctrl. **b**, Time course of normalized $^{86}\text{Rb}^+$ uptake into liposomes with reconstituted KtrAB transporter (with ATP or ADP) or KtrB alone (with ATP). Control measurement was with liposomes formed in the presence of KtrA alone; normalization after addition of valinomycin at end of assay. Mean \pm s.e.m. calculated from six assays corresponding to triplicates from two separate liposome preparations. c.p.m., counts per minute.

structure and the KtrA structure that forms the KtrAB-ATP complex (Supplementary Fig. 3a). This suggests that unlike truncated KtrA (composed by N lobe alone; Supplementary Fig. 9) where ligand binding and ring conformation are uncoupled²¹, in the full-length KtrA the ATP molecule induces the formation of a unique set of intradimer interactions which determine the conformation of the octameric ring. The absence of the γ -phosphate in ADP is reflected in the lack of ligand-mediated intra-dimer interactions and in the relaxation of the ring.

Implications for KtrAB activation

We have uncovered several differences between KtrAB and the BK^{13–15} and MthK^{11,12,24} channels which have an impact on their mechanisms of transport activation. In channels, the RCK rings are covalently tethered to the C terminus of a single channel pore domain and the rings expand symmetrically upon binding of an activating ligand. In KtrAB, the cytoplasmic loops of the KtrB homodimer mediate the interaction with the KtrA (RCK) ring and the ring contracts asymmetrically upon binding of ATP.

To propose an activation mechanism for KtrAB we need to define its 'gates' (that is, regions that undergo a conformational change and close/open the ion pathway)—ion transporters²⁵ and channels²⁶ can have multiple gates. In KtrAB, the intramembrane loop has been proposed to be a gate^{9,19}. A comparison between KtrB and TrkH structures (Supplementary Fig. 7a–d) raises the possibility of another gate region; it shows the M2b helix/D1–D2 loop region adopting different conformations. In TrkH, helices M2a and M2b (in repeat D1) are continuous and the ordered amino acid stretch of the D1–D2 loop is positioned so that it partially obstructs the cytoplasmic ion pathway of KtrB. This indicates that this region is able to adopt different positions and could function as a gate.

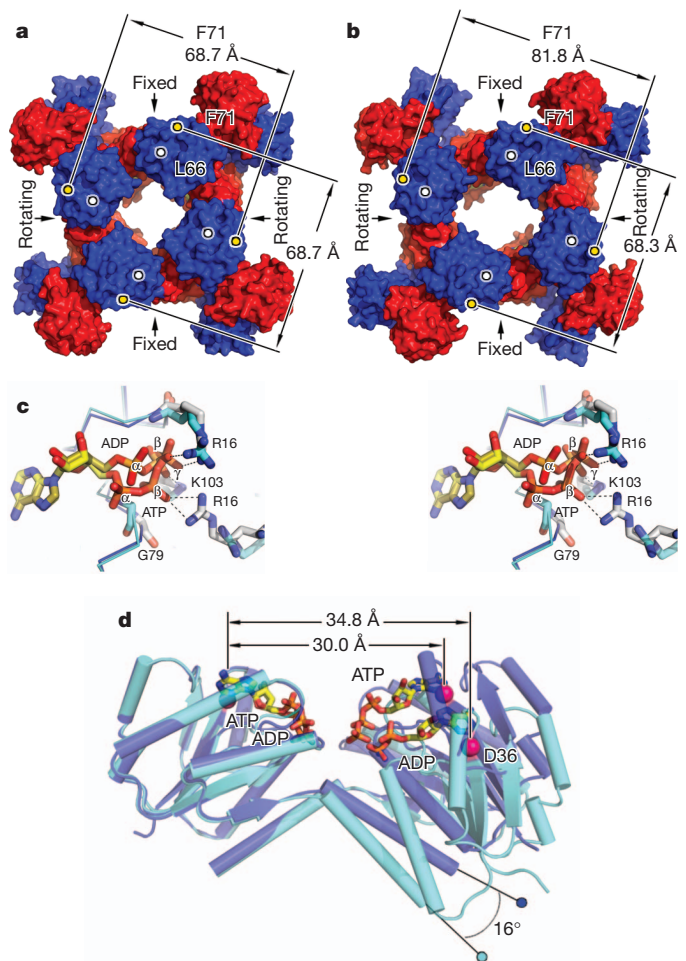


Figure 4 | Conformational changes in the KtrA octameric ring. **a**, **b**, Surface representations of isolated KtrA-ATP ring (**a**) and KtrA-ADP ring (**b**). KtrA subunits are shown in red and blue. F71 C α is shown as yellow dots; L66 C α is shown as white dots. Dimer-to-dimer interfaces that undergo a rigid-body rotation between ATP- and ADP-bound states are indicated as rotating; the other dimer-to-dimer interfaces are indicated as fixed. **c**, Stereo view of superposed ligand-binding sites from KtrA-ADP (light blue C α trace and side-chain stick) and KtrA-ATP (dark blue C α trace and white side-chain stick). Residues, ATP and ADP molecules and ligand phosphates are labelled. Some ligand-protein interactions are shown as dashed lines. **d**, Superposition of a KtrA-ATP (dark blue) and KtrA-ADP (light blue) dimer through one of the subunits; only N lobes are shown. Ligands are shown as stick representation. The distance separating D36 residues (red spheres) in dimer subunits and dimer hinge angle change are indicated.

We foresee therefore two possible activation mechanisms for KtrAB (see also Supplementary Discussion^{27,28}). For model 1, the asymmetric KtrA contraction/expansion occurs along the diagonal defined by the lateral contacts (Fig. 5a, b). The direct connection between the KtrB C terminus, involved in the lateral-contact region, and the R417/intramembrane loop gate suggests that this would be the KtrA controlled gate. There are two pieces of evidence against this model: first, the R417/intramembrane gate seems to be closed in the KtrAB structure bound to ATP, an activating ligand. Second, sequential truncations of 1–3 residues in the C terminus have no apparent functional consequence. For model 2, the asymmetric change in KtrA occurs along the diagonal defined by the tip contacts (Fig. 5a, c). We see no obvious direct structural link between the tip-contact region, formed by the D1–D2 loop in KtrB, and the R417/intramembrane gate (Fig. 2a). A clear alternative is that the D1–D2 loop is the KtrA controlled gate, whereas the R417/intramembrane gate is activated by an as-yet-unknown stimulus. Upon ATP binding, the D1–D2 loop

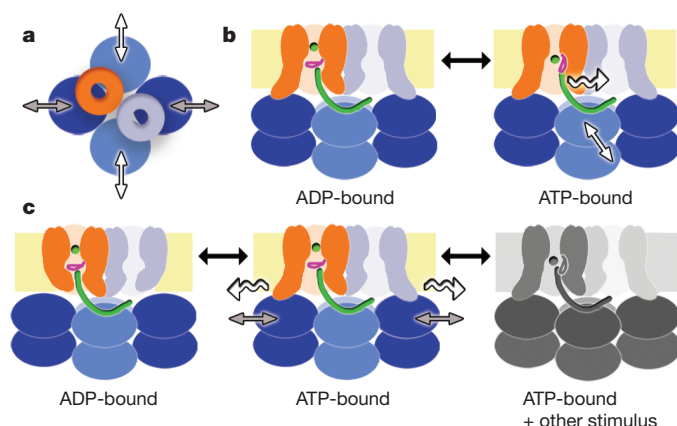


Figure 5 | Models for KtrAB activation. Cartoon with the KtrB homodimer in orange and grey; KtrA subunits as dark- and light-blue ovals; intramembrane loop in magenta; C terminus as green coil; and K^+ as green sphere. **a**, Extracellular view with asymmetric conformational changes in the KtrA ring occurring along two possible directions: along lateral contact (pair of white arrows) or along tip contact (grey arrows). **b**, **c**, Side views of KtrAB. For model 1 (**b**), KtrA asymmetrical conformational change along lateral contact is propagated through the KtrB C terminus to intramembrane loop gate (indicated by wavy arrow). For model 2 (**c**), KtrA asymmetrical conformational change along tip contact alters conformation of the KtrB D1–D2 loop gate (wavy arrows), opening a gate. For full activation the intramembrane loop gate has to open upon an as-yet-unknown stimulus (cartoon representation in dark grey).

would adopt the conformation seen in the KtrAB structure, opening the cytoplasmic pore and resulting in an increase in transporter activity, as detected by our functional assay. Full transporter activity, as seen in cells⁸, would require not just ATP but also the influence of another stimulus to activate the intramembrane gate.

In summary, this study provides novel insights and raises new ideas about the structure and mechanism of regulation of the KtrAB K^+ transporter, offering a better understanding of the function of the Trk/HKT transporters in bacteria, archaea, fungi and plants.

METHODS SUMMARY

Briefly, *B. subtilis* KtrB was overexpressed in the BL21 (DE3) *E. coli* strain, protein extracted with dodecylmaltoside, affinity purified in cobalt matrix and further purified by size-exclusion chromatography. Tag-less *B. subtilis* KtrA protein was overexpressed in BL21 (DE3), and protein was purified by affinity chromatography in an ADP-agarose matrix. KtrAB complex was assembled by mixing KtrB and KtrA. Preparation of proteoliposomes and $^{86}Rb^+$ flux assay was performed as previously described with some variations^{6,29}. For crystallization the complex was further purified and detergent exchanged to Cymal6 by size-exclusion chromatography. Structure determination and representation made use of XD3³⁰, MOSFLM³¹, PHARE³², PHASER³³, PHENIX³⁴, RAVE³⁵, COOT³⁶, Diffraction Anisotropy Server³⁷, PYMOL³⁸, STRAP³⁹, ESPRIT⁴⁰, Hollow⁴¹ and CCP4⁴² programs. Growth complementation assay followed protocols previously described^{6,8}. Protein quality control was performed by mass spectrometry.

Full Methods and any associated references are available in the online version of the paper.

Received 4 October 2012; accepted 5 March 2013.

- Corratgé-Faillie, C. *et al.* Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. *Cell. Mol. Life Sci.* **67**, 2511–2532 (2010).
- Durell, S. R., Hao, Y., Nakamura, T., Bakker, E. P. & Guy, H. R. Evolutionary relationship between K^+ channels and symporters. *Biophys. J.* **77**, 775–788 (1999).
- Nakamura, T., Yuda, R., Unemoto, T. & Bakker, E. P. KtrAB, a new type of bacterial K^+ -uptake system from *Vibrio alginolyticus*. *J. Bacteriol.* **180**, 3491–3494 (1998).
- Hänelt, I. *et al.* KtrB, a member of the superfamily of K^+ transporters. *Eur. J. Cell Biol.* **90**, 696–704 (2011).
- Tholema, N., Bakker, E. P., Suzuki, A. & Nakamura, T. Change to alanine of one out of four selectivity filter glycines in KtrB causes a two orders of magnitude decrease in

- the affinities for both K^+ and Na^+ of the Na^+ dependent K^+ uptake system KtrAB from *Vibrio alginolyticus*. *FEBS Lett.* **450**, 217–220 (1999).
- Albright, R. A., Joh, K. & Morais-Cabral, J. H. Probing the structure of the dimeric KtrB membrane protein. *J. Biol. Chem.* **282**, 35046–35055 (2007).
 - Maser, P. *et al.* Glycine residues in potassium channel-like selectivity filters determine potassium selectivity in four-loop-per-subunit HKT transporters from plants. *Proc. Natl Acad. Sci. USA* **99**, 6428–6433 (2002).
 - Tholema, N. *et al.* All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K^+ uptake by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.* **280**, 41146–41154 (2005).
 - Cao, Y. *et al.* Crystal structure of a potassium ion transporter, TrkH. *Nature* **471**, 336–340 (2011).
 - Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
 - Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
 - Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
 - Wu, Y., Yang, Y., Ye, S. & Jiang, Y. Structure of the gating ring from the human large-conductance Ca^{2+} -gated K^+ channel. *Nature* **466**, 393–397 (2010).
 - Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y. & MacKinnon, R. Structure of the human BK channel Ca^{2+} -activation apparatus at 3.0 Å resolution. *Science* **329**, 182–186 (2010).
 - Yuan, P., Leonetti, M. D., Hsiung, Y. & MacKinnon, R. Open structure of the Ca^{2+} gating ring in the high-conductance Ca^{2+} -activated K^+ channel. *Nature* **481**, 94–97 (2012).
 - Holtmann, G., Bakker, E. P., Uozumi, N. & Bremer, E. KtrAB and KtrCD: two K^+ uptake systems in *Bacillus subtilis* and their role in adaptation to hypertonicity. *J. Bacteriol.* **185**, 1289–1298 (2003).
 - Berry, S. *et al.* Potassium uptake in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803 mainly depends on a Ktr-like system encoded by slr1509 (ntpJ). *FEBS Lett.* **548**, 53–58 (2003).
 - Matsuda, N. *et al.* Na^+ -dependent K^+ uptake Ktr system from the cyanobacterium *Synechocystis* sp. PCC 6803 and its role in the early phases of cell adaptation to hyperosmotic shock. *J. Biol. Chem.* **279**, 54952–54962 (2004).
 - Hanelt, I. *et al.* Gain of function mutations in membrane region M2C2 of KtrB open a gate controlling K^+ transport by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.* **285**, 10318–10327 (2010).
 - Kroning, N. *et al.* ATP binding to the KTN/RCK subunit KtrA from the K^+ -uptake system KtrAB of *Vibrio alginolyticus*: its role in the formation of the KtrAB complex and its requirement *in vivo*. *J. Biol. Chem.* **282**, 14018–14027 (2007).
 - Albright, R. A., Ibar, J. L., Kim, C. U., Gruner, S. M. & Morais-Cabral, J. H. The RCK domain of the KtrAB K^+ transporter: multiple conformations of an octameric ring. *Cell* **126**, 1147–1159 (2006).
 - Roosild, T. P., Miller, S., Booth, I. R. & Choe, S. A mechanism of regulating transmembrane potassium flux through a ligand-mediated conformational switch. *Cell* **109**, 781–791 (2002).
 - Heginbotham, L., Kolmakova-Partensky, L. & Miller, C. Functional reconstitution of a prokaryotic K^+ channel. *J. Gen. Physiol.* **111**, 741–749 (1998).
 - Ye, S., Li, Y., Chen, L. & Jiang, Y. Crystal structures of a ligand-free MthK gating ring: insights into the ligand gating mechanism of K^+ channels. *Cell* **126**, 1161–1173 (2006).
 - Olesen, C. *et al.* The structural basis of calcium transport by the calcium pump. *Nature* **450**, 1036–1042 (2007).
 - Cuello, L. G. *et al.* Structural basis for the coupling between activation and inactivation gates in K^+ channels. *Nature* **466**, 272–275 (2010).
 - Trudeau, M. C. Unlocking the mechanisms of HCN channel gating with locked-open and locked-closed channels. *J. Gen. Physiol.* **140**, 457–461 (2012).
 - Hilf, R. J. & Dutzler, R. A prokaryotic perspective on pentameric ligand-gated ion channel structure. *Curr. Opin. Struct. Biol.* **19**, 418–424 (2009).
 - Nimigean, C. M. A radioactive uptake assay to measure ion transport across ion channel-containing liposomes. *Nature Protocols* **1**, 1207–1212 (2006).
 - Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
 - Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
 - Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
 - McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
 - Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
 - Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55**, 941–944 (1999).
 - Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
 - Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
 - Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1 (2010).
 - Gille, C. & Frommel, C. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17**, 377–378 (2001).
 - Gouet, P., Robert, X. & Courcelle, E. ESPRIT/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
 - Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).

42. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful for access to ID14-1/ID14-4/ID-29 at ESRF (through the Portuguese BAG), PXII at SLS, XRD1 at ELETTRA and PROXIMA1 at SOLEIL and thank the respective support staff. A.S. was supported by FEBS (Long term fellowship). This work was funded by EMBO (Installation grant), by FEDER funds through the Operational Competitiveness Program–COMPETE and by National Funds through FCT–Fundação para a Ciência e a Tecnologia under the projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011), FCOMP-01-0124-FEDER-009028 (PTDC/BIA-PRO/099861/2008) and FCOMP-01-0124-FEDER-010781 (PTDC/QUI-BIQ/105342/2008). We also thank

G. Gabant and M. Cadene at the ‘Plateforme de Spectrometrie de Masse’ at CBM, CNRS, Orleans for mass spectrometry analysis, and C. Harley for critical reading of the manuscript.

Author Contributions R.S.V.-P. performed all crystallographic work with the support of J.H.M.-C.; R.S.V.-P. and A.S. performed the functional and biochemical characterization; J.H.M.-C., R.S.V.-P. and A.S. designed experiments and wrote the manuscript.

Author Information Structures have been deposited in the Protein Data Bank under accessions 4J7C (KtrAB complex), 4J90 (KtrA–ATP) and 4J91 (KtrA–ADP). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.M.-C. (jcabral@ibmc.up.pt).

METHODS

Protein expression and purification. N-terminal His-tagged KtrB was overexpressed in *E. coli* BL21 (DE3), in LB media at 37 °C for 2.5 h after IPTG induction. After cell lysis in Buffer A (50 mM Tris-HCl pH 8.0, 120 mM NaCl, 30 mM KCl) supplemented with protease inhibitors, the lysate was supplemented with 40 mM DDM (*n*-dodecyl- β -D-maltoside, solgrade from Anatrace) for membrane protein extraction overnight at 4 °C. Spin-cleared lysate was loaded onto a Talon metal affinity resin (Clontech) and washed with Buffer B (50 mM Tris-HCl pH 8.0, 120 mM NaCl, 30 mM KCl, 1 mM DDM). Protein was eluted with Buffer B supplemented with 150 mM imidazole. Eluted fractions were concentrated to ~ 2.5 mg ml⁻¹ and incubated overnight at 4 °C with thrombin for tag cleavage. At this point the protein was either directly used for KtrAB complex assembly or further purified by size exclusion with a Superdex-200 column.

Tag-less KtrA was overexpressed in *E. coli* BL21 (DE3), in LB media at 20 °C for 14–16 h after IPTG induction. Cells lysis was done in Buffer C (50 mM Tris-HCl pH 7.5, 50 mM KCl, 5 mM DTT) and cleared lysate loaded into an anion exchange column. Fractions containing KtrA were incubated with ADP-agarose resin (Innova Biosciences) overnight at 4 °C. Beads were washed thoroughly with Buffer D (50 mM Tris-HCl, 150 mM KCl, 1 mM TCEP) and protein was eluted in the same buffer supplemented with 5 mM of adenosine-containing nucleotide. At this point the protein was concentrated to ~ 3 mg ml⁻¹ and either directly used for KtrAB complex assembly or further purified by size exclusion with a Superdex-S200 column.

KtrAB complex was assembled by mixing individually purified KtrB and KtrA batches, at a pre-defined ratio. The complex mixture was further concentrated (~ 20 mg ml⁻¹) and purified in a size-exclusion Superdex-200 column with Buffer E (20 mM Tris-HCl pH 8.0, 120 mM NaCl, 30 mM KCl, 5 mM DTT, 1.5 mM 6-cyclohexyl-1-hexyl- β -D-maltoside (cymal-6, anagrade from Anatrace)).

Crystallization. KtrAB complex purified in Buffer E supplemented with 1 mM ATP (ATP was concentrated to ~ 10 mg ml⁻¹ and dialysed overnight against the same buffer before crystallization assays. KtrAB complex crystals were obtained by the sitting-drop vapour diffusion method, mixing 1 μ l of protein with 1 μ l of crystallization solution. The best diffracting crystals grew in 100 mM *N*-(2-acetamido)-iminodiacetic acid (ADA) pH 6.5, 20% polyethylene glycol 400 (PEG 400) and 200 mM ammonium sulphate at 20 °C. Crystals were flash-frozen in liquid nitrogen after addition of mother liquor with 45% PEG 400 directly to the crystal drop.

Before crystallization KtrA-ATP and KtrA-ADP were run in a size-exclusion chromatography column equilibrated with Buffer F (50 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM DTT, 1 mM ATP or ADP). KtrA-ATP crystals were grown by sitting-drop vapour diffusion with 100 mM HEPES pH 7.5, 3% polyethylene glycol 6000 (PEG 6000) and 2.5% 2-methyl-2,4-pentanediol (MPD) at 20 °C. Crystals were flash frozen in liquid nitrogen after being transferred to identical mother liquor containing 6% PEG 6000 and 30% glycerol. KtrA-ADP crystals were obtained with 100 mM citrate pH 5.6, 14.5% PEG 2000 and 200 mM ammonium sulphate at 20 °C. Crystals were flash-frozen in liquid nitrogen after addition of modified mother liquor, with 15% PEG 2000 and 20% glycerol, directly to the crystal drop.

Data processing and structure determination. Diffraction data were processed with XDS³⁰ or MOSFLM³¹, followed by scaling with CCP4 package⁴². Data from KtrAB crystals is highly anisotropic and was analysed, truncated and sharpened in the Diffraction Anisotropy Server³⁷ (<http://services.mbi.ucla.edu/anisotscale/>). Recommended resolution limits after anisotropic correction were 3.8 Å, 4.1 Å and 3.3 Å, respectively, along the *a**, *b** and *c** reciprocal space directions.

KtrA-ATP structure was determined by molecular replacement with PHASER³³ using as search models the truncated KtrA N-lobe structure (Protein Data Bank code 2HMU) and C-lobe homology models generated by PHYRE³². Improvement of electron-density maps, in particular on the C-lobe region, was achieved by cross-crystal averaging between KtrA-ATP and truncated KtrA using MAMA³⁵ and DMmulti (CCP4 package). The structure was finalized with cycles of model building with COOT³⁶ and refinement with PHENIX³⁴ and REFMAC (CCP4 package). The two subunits in the asymmetric unit include residues 6 to 222. Ramachandran plot for KtrA-ATP structure contains: 0% outliers, 1.63% allowed, 98.37% preferred.

The N and C lobes of KtrA-ATP structure were used for molecular replacement with the KtrA-ADP data. Cycles of averaging between the four subunits of KtrA-ADP and the two subunits of KtrA-ATP improved maps. Model building and refinement produced a final model of KtrA-ADP that includes: in chain A, residues 8–157, 162–174 (built as alanines), 177–222; in chain B, residues 8–222; in chain C, residues 8–147, 152–157, 161–176 (built as alanines), 177–222; in chain D, residues 8–187, 188–192 (built as alanines), 196–222. Ramachandran plot for KtrA-ADP structure contains: 0% outliers, 1.82% allowed, 98.18% preferred.

The KtrAB structure was determined by molecular replacement using as search models a poly-alanine version of the TrkH dimer structure (Protein Data Bank

code 3PJZ) and the KtrA-ATP octameric ring structure. Cycles of four-fold averaging for the KtrB subunits and eight-fold averaging for the KtrA ring subunits together with cross-crystal averaging with the KtrA-ATP structure resulted in great improvements in the electron-density maps and allowed tracing of the KtrB chain. Chain trace was confirmed with a selenomethionine anomalous map calculated to 7.4 Å, which at a 3.6 σ contour revealed the position of 9 methionines out of 14 possible (Supplementary Fig. 10), and through a comparison with the TrkH structure. Structure was finalized through cycles of model building and refinement. The model was refined against data to 3.5 Å as there was no visible improvement in refinement statistics or map quality with data to the recommended limit (see above). In the KtrAB structure the KtrB subunits span residues 15 to 445 (except for a missing stretch formed by residues 103 and 104) whereas the KtrA subunits span residues 7–222. Ramachandran plot for KtrAB crystals: 0% outliers, 5.94% allowed, 94.06% preferred. Model analysis was performed with PYMOL³⁸, STRAP³⁹, ESPrit⁴⁰ and Hollow⁴¹.

Complementation assay. The complementation assay with TK2420 follows procedures previously described^{3,6,8}. Briefly, mutants and wild-type KtrB and KtrA were cloned into a dicistronic constitutive expression plasmid. The TK2420 strain was transformed with the different plasmids and plated on LBK (Luria-Bertani broth where NaCl is replaced by KCl) agar plates. Single colonies were picked and grown in liquid LBK overnight at 37 °C. A 2-ml aliquot of cells was gently pelleted and re-suspended in 400 μ l of minimal media containing 0.1 mM K⁺. A 15 μ l aliquot of re-suspended cells was then used to inoculate 5 ml of minimal media prepared with different K⁺ concentrations (0.1, 0.3, 1, 2, 6, 10, 30 or 115 mM) and grown for 15 h at 37 °C at which point optical density at 595 nm was measured.

Preparation of proteoliposomes. Preparation follows previously described methods^{6,29}. Polar *E. coli* lipids in chloroform were dried under a stream of argon, washed in pentane and dried again. Residual solvent was removed under vacuum. Lipids were re-suspended at 10 mg ml⁻¹ in swelling buffer (150 mM KCl, 10 mM HEPES, 5 mM *N*-methylglucamine, pH 7.6) using a bath sonicator. Lipids were solubilized by adding decylmaltoside (powder) to 40 mM. After 2 h of gentle agitation, dodecylmaltoside-solubilized protein, in KtrB buffer (150 mM KCl, 20 mM Tris-HCl, 5 mM DTT, 0.5 mM dodecylmaltoside, 200 μ M NTP, pH 8.0), was added at 1:100 w:w protein-to-lipid ratio and incubated for 30 min at room temperature. KtrAB complex was assembled just before reconstitution with excess of either ATP-bound KtrA or ADP-bound KtrA. The amount of protein added to 100 μ l solubilized lipids was adjusted to contain 10 μ g KtrB both when reconstituted alone or as KtrAB complex. Detergent was removed by dialysis against swelling buffer supplemented with 20 μ M ATP or ADP and with Na⁺ concentration at 40 μ M, using 15 kDa cutoff dialysis membranes for 4 days at 4 °C. Control liposomes were prepared similarly, except that only KtrA was added to the solubilized lipids.

⁸⁶Rb⁺ flux assay. Assay as previously described^{6,29}, with small modifications. A gradient of K⁺ across the liposome membrane was generated, immediately before the assay, by loading 100 μ l aliquot of liposomes into a 5 ml disposable spin column containing Sephadex G50 (fine) pre-swollen with sorbitol buffer (20 μ M KCl, 150 mM sorbitol, 10 mM HEPES, 5 mM *N*-methylglucamine, pH 7.6) and centrifuging for 1 min at 700g. After forming the gradient, liposomes were supplemented with 20 μ M ADP or ATP (final concentration of Na⁺ was set to 40 μ M) and mixed with three volume of ⁸⁶Rb⁺ assay buffer (⁸⁶Rb⁺ at $\sim 2,000$ counts μ l⁻¹ in sorbitol buffer). At different time points 100 μ l of uptake reaction were loaded on a 1.5-ml bed volume Dowex cation exchange columns (pre-washed in 150 mM sorbitol solution plus 5 mg ml⁻¹ BSA to avoid unspecific binding to resin, and then pre-equilibrated in 150 mM sorbitol solution) and eluted with 6% sorbitol solution. The eluate was collected in 4-ml scintillation vials and mixed with 1.5 ml Optiphase scintillation fluid. At the end of the experiment 900 nM valinomycin was added to the last 100 μ l aliquot of the uptake reaction, incubated for 3 min and then loaded on Dowex columns as described above. The uptake of ⁸⁶Rb⁺ into liposomes was expressed as the percentage of the counts measured after valinomycin treatment (this provides a correction for the internal volume of liposomes in the reaction sample). Valinomycin was also applied at earlier time points to check whether the long timescale of the experiments could affect the final counts, but the values were similar to those measured after 2 h. For all conditions means \pm s.e.m. of triplicates were calculated for two separate liposome preparations.

KtrA ligand binding. Potential ligands for KtrA were screened by making use of the tight binding of isolated KtrA to an ADP-agarose matrix and by evaluating which small molecules elute the protein from the beads. Purified KtrA was incubated with the beads at 4 °C. Beads were washed with buffer and then incubated for 30 min with 1 mM of different ligands: ATP (adenosine triphosphate), ADP (adenosine diphosphate), AMP (adenosine monophosphate), FAD (flavin mononucleotide), NAD (nicotinamide adenine dinucleotide), NADH (reduced NAD), NADPH (reduced nicotinamide adenine dinucleotide phosphate) and GTP (guanosine triphosphate). Eluted fractions were run in an SDS-PAGE.

A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34

Dominik A. Riechers^{1,2}, C. M. Bradford^{1,3}, D. L. Clements⁴, C. D. Dowell^{1,3}, I. Pérez-Fournon^{5,6}, R. J. Ivison^{7,8}, C. Bridge¹, A. Conley⁹, Hai Fu¹⁰, J. D. Vieira¹, J. Wardlow¹⁰, J. Calanog¹⁰, A. Cooray^{1,10}, P. Hurley¹¹, R. Neri¹², J. Kamenetzky¹³, J. E. Aguirre¹⁴, B. Altieri¹⁵, V. Arumugam⁸, D. J. Benford¹⁶, M. Béthermin^{17,18}, J. Bock^{1,3}, D. Burgarella¹⁹, A. Cabrera-Lavers^{5,6,20}, S. C. Chapman²¹, P. Cox¹², J. S. Dunlop⁸, L. Earle⁹, D. Farrah²², P. Ferrero^{5,6}, A. Franceschini²³, R. Gavazzi²⁴, J. Glenn^{9,13}, E. A. Gonzalez Solares²¹, M. A. Gurwell²⁵, M. Halpern²⁶, E. Hatziminaoglou²⁷, A. Hyde⁴, E. Ibar⁷, A. Kovács^{1,28}, M. Krips¹², R. E. Lupu¹⁴, P. R. Maloney⁹, P. Martinez-Navajas^{5,6}, H. Matsuhara²⁹, E. J. Murphy^{1,30}, B. J. Naylor³, H. T. Nguyen^{1,3}, S. J. Oliver¹¹, A. Omont²⁴, M. J. Page³¹, G. Petitpas²⁵, N. Rangwala¹³, I. G. Roseboom^{8,11}, D. Scott²⁶, A. J. Smith¹¹, J. G. Staguhn^{16,32}, A. Streblyanska^{5,6}, A. P. Thomson⁸, I. Valtchanov¹⁵, M. Viero¹, L. Wang¹¹, M. Zemcov^{1,3} & J. Zmuidzinas^{1,3}

Massive present-day early-type (elliptical and lenticular) galaxies probably gained the bulk of their stellar mass and heavy elements through intense, dust-enshrouded starbursts—that is, increased rates of star formation—in the most massive dark-matter haloes at early epochs. However, it remains unknown how soon after the Big Bang massive starburst progenitors exist. The measured redshift (z) distribution of dusty, massive starbursts has long been suspected to be biased low in z owing to selection effects¹, as confirmed by recent findings of systems with redshifts as high as ~ 5 (refs 2–4). Here we report the identification of a massive starburst galaxy at $z = 6.34$ through a submillimetre colour-selection technique. We unambiguously determined the redshift from a suite of molecular and atomic fine-structure cooling lines. These measurements reveal a hundred billion solar masses of highly excited, chemically evolved interstellar medium in this galaxy, which constitutes at least 40 per cent of the baryonic mass. A ‘maximum starburst’ converts the gas into stars at a rate more than 2,000 times that of the Milky Way, a rate among the highest observed at any epoch. Despite the overall downturn in cosmic star formation towards the highest redshifts⁵, it seems that environments mature enough to form the most massive, intense starbursts existed at least as early as 880 million years after the Big Bang.

We have searched 21 deg² of the Herschel/SPIRE data of the HerMES blank field survey⁶ at wavelengths 250–500 μm for ‘ultra-red’ sources with flux densities $S_{250\mu\text{m}} < S_{350\mu\text{m}} < S_{500\mu\text{m}}$ and $S_{500\mu\text{m}}/S_{350\mu\text{m}} > 1.3$, that is, galaxies that are significantly redder (and thus, potentially at higher redshift) than massive starbursts discovered thus far. This selection yields five candidate ultra-red sources down to a flux limit of 30 mJy at 500 μm ($>5\sigma$ and above the confusion noise; see Supplementary Information section 1 for additional details), corresponding to a source density of $\leq 0.24 \text{ deg}^{-2}$. For comparison, models

of number counts in the Herschel/SPIRE bands suggest a space density of massive starburst galaxies at $z > 6$ with $S_{500\mu\text{m}} > 30 \text{ mJy}$ of 0.014 deg^{-2} (ref. 7).

To understand the nature of galaxies selected by this technique, we have obtained full frequency scans of the 3-mm and 1-mm bands towards HFLS 3 (also known as 1HERMES S350 J170647.8+584623; $S_{500\mu\text{m}}/S_{350\mu\text{m}} = 1.45$), the brightest candidate discovered in our study. These observations, augmented by selected follow-up over a broader wavelength range, unambiguously determine the galaxy redshift to be $z = 6.3369 \pm 0.0009$ based on a suite of 7 CO lines, 7 H₂O lines, and OH, OH⁺, H₂O⁺, NH₃, [C I] and [C II] lines detected in emission and absorption (Fig. 1). At this redshift, the Universe was just 880 million years old (or one-sixteenth of its present age), and 1'' on the sky corresponds to a physical scale of 5.6 kpc. Further observations from optical to radio wavelengths reveal strong continuum emission over virtually the entire wavelength range between 2.2 μm and 20 cm, with no detected emission shortward of 1 μm (see Supplementary Information section 2 and Supplementary Figs 1–11 for additional details).

HFLS 3 hosts an intense starburst. The 870- μm flux of HFLS 3 is >3.5 times higher than those of the brightest high-redshift starbursts in a 0.25-deg² region containing the Hubble Ultra Deep Field (HUDF)⁸. From the continuum spectral energy distribution (Fig. 2), we find that the far-infrared (FIR) luminosity L_{FIR} and inferred star formation rate (SFR) of 2,900 $M_{\text{sun}} \text{ yr}^{-1}$ of HFLS 3 (where M_{sun} is the solar mass) are 15–20 times those of the prototypical local ultra-luminous starburst Arp 220, and $>2,000$ times those of the Milky Way (Table 1 and Supplementary Information section 3). The SFR of HFLS 3 alone corresponds to ~ 4.5 times the ultraviolet-based SFR of all $z = 5.5$ – 6.5 star-forming galaxies in the HUDF combined⁹, but the rarity and dust obscuration of ultra-red sources like HFLS 3 implies that they do not dominate the ultraviolet photon density needed to reionize the Universe¹⁰.

¹California Institute of Technology, 1200 East California Boulevard, MC 249-17, Pasadena, California 91125, USA. ²Cornell University, 220 Space Sciences Building, Ithaca, New York 14853, USA. ³Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA. ⁴Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK. ⁵Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain. ⁶Departamento de Astrofísica, Universidad de La Laguna, E-38205 La Laguna, Tenerife, Spain. ⁷UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK. ⁸Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK. ⁹Center for Astrophysics and Space Astronomy 389-UCB, University of Colorado, Boulder, Colorado 80309, USA. ¹⁰Department of Physics and Astronomy, University of California, Irvine, California 92697, USA. ¹¹Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK. ¹²Institut de RadioAstronomie Millimétrique, 300 Rue de la Piscine, Domaine Universitaire, F-38406 Saint Martin d'Hères, France. ¹³Department of Astrophysical and Planetary Sciences, CASA 389-UCB, University of Colorado, Boulder, Colorado 80309, USA. ¹⁴Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁵Herschel Science Centre, European Space Astronomy Centre, Villanueva de la Cañada, 28691 Madrid, Spain. ¹⁶Observational Cosmology Laboratory, Code 665, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹⁷Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu – CNRS – Université Paris Diderot, CEA Saclay, Point courrier 131, F-91191 Gif-sur-Yvette, France. ¹⁸Institut d'Astrophysique Spatiale (IAS), Bâtiment 121, Université Paris-Sud 11 and CNRS, UMR 8617, F-91405 Orsay, France. ¹⁹Aix-Marseille Université, CNRS, Laboratoire d'Astrophysique de Marseille, UMR7326, F-13388 Marseille, France. ²⁰Grantecan S.A., Centro de Astrofísica de La Palma, Cuesta de San Jose, E-38712 Brea Baja, La Palma, Spain. ²¹Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. ²²Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA. ²³Dipartimento di Fisica e Astronomia, Università di Padova, vicolo Osservatorio 3, I-35122 Padova, Italy. ²⁴Institut d'Astrophysique de Paris, UMR 7095, CNRS, UPMC Univ. Paris 06, 98bis Boulevard Arago, F-75014 Paris, France. ²⁵Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ²⁶Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ²⁷ESO, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany. ²⁸Institute for Astrophysics, University of Minnesota, 116 Church Street Southeast, Minneapolis, Minnesota 55455, USA. ²⁹Institute for Space and Astronautical Science, Japan Aerospace and Exploration Agency, Sagamihara, Kanagawa 229-8510, Japan. ³⁰Infrared Processing and Analysis Center, MS 100-22, California Institute of Technology, Pasadena, California 91125, USA. ³¹Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking RH5 6NT, UK. ³²Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA.

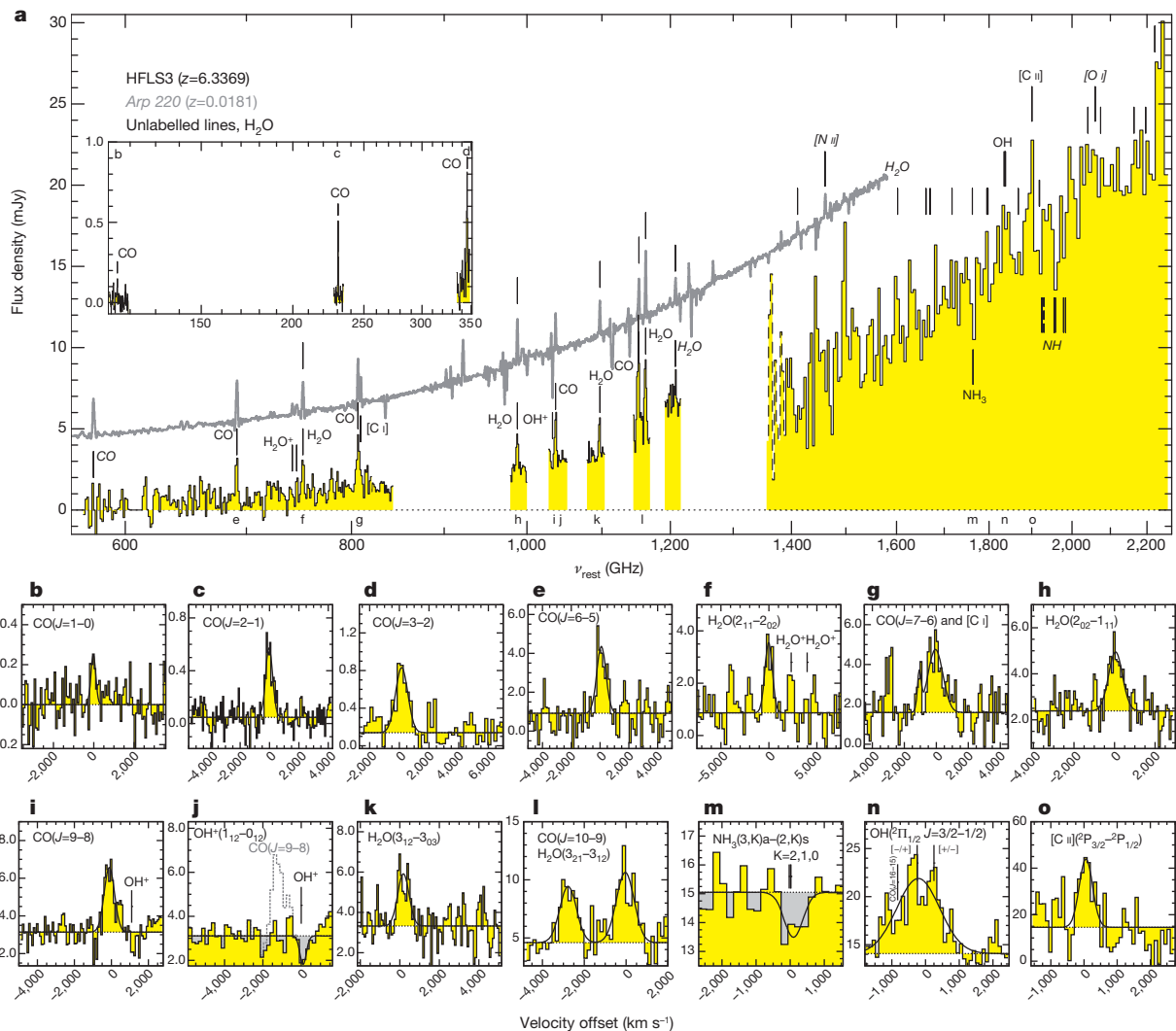


Figure 1 | Redshift identification through molecular and atomic spectroscopy of HFLS 3. **a**, Black trace, wide-band spectroscopy in the observed-frame 19–0.95-mm (histogram; rest-frame 2,600–130 μ m) wavelength range with CARMA (3 mm; ‘blind’ frequency scan of the full band), the PdBI (2 mm), the JVLA (19–6 mm) and CSO/Z-spec (1 mm; instantaneous coverage). (CARMA, Combined Array for Research in Millimeter-wave Astronomy; PdBI, Plateau de Bure Interferometer; JVLA, Jansky Very Large Array; and CSO, Caltech Submillimeter Observatory.) This uniquely determines the redshift of HFLS 3 to be $z = 6.3369$ based on the detection of a series of H_2O , CO, OH, OH^+ , NH_3 , [C I] and [C II] emission and absorption lines. **b–o**, Detailed profiles of detected lines (histograms; rest frequencies are indicated by corresponding letters in **a**). 1-mm lines (**m–o**) are deeper, interferometric confirmation observations for NH_3 , OH (both PdBI) and [C II]

HFLS 3 is a massive, gas-rich galaxy. From the spectral energy distribution and the intensity of the CO and [C II] emission, we find a dust mass of $M_d = 1.3 \times 10^9 M_{\text{sun}}$ and total molecular and atomic gas masses of respectively $M_{\text{gas}} = 1.0 \times 10^{11} M_{\text{sun}}$ and $M_{\text{H I}} = 2.0 \times 10^{10} M_{\text{sun}}$. These masses are 15–20 times those of Arp 220, and correspond to a gas-to-dust ratio of ~ 80 and a gas depletion timescale of $M_{\text{gas}}/\text{SFR} \approx 36$ Myr. These values are comparable to lower-redshift submillimetre-selected starbursts^{11,12}. From the [C I] luminosity, we find an atomic carbon mass of $4.5 \times 10^7 M_{\text{sun}}$. At the current SFR of HFLS 3, this level of carbon enrichment could have been achieved through supernovae on a timescale of $\sim 10^7$ yr (ref. 13). The profiles of the molecular and atomic emission lines typically show two velocity components (Fig. 1 and Supplementary Figs 5 and 7). The gas is distributed over a region of 1.7 kpc radius with a high velocity gradient and dispersion (Fig. 3). This suggests a dispersion-dominated galaxy with a dynamical mass of $M_{\text{dyn}} = 2.7 \times 10^{11} M_{\text{sun}}$. The

(CARMA) not shown in **a**. The line profiles are typically asymmetric relative to single Gaussian fits, indicating the presence of two principal velocity components at redshifts of 6.3335 and 6.3427. The implied CO, [C I] and [C II] line luminosities are respectively $(5.08 \pm 0.45) \times 10^6 L_{\text{sun}}$, $(3.0 \pm 1.9) \times 10^8 L_{\text{sun}}$ and $(1.55 \pm 0.32) \times 10^{10} L_{\text{sun}}$. Strong rest-frame submillimetre to FIR continuum emission is detected over virtually the entire wavelength range. For comparison, the Herschel/SPIRE spectrum of the nearby ultra-luminous infrared galaxy Arp 220²⁰ is overplotted in grey (**a**). Lines labelled in *italic* are tentative detections or upper limits (see Supplementary Table 2). Most of the bright spectral features detected in Arp 220^{20,21} are also detected in HFLS 3 (in spectral regions not blocked by the terrestrial atmosphere). See Supplementary Information sections 2–4 for more details.

gas mass fraction in galaxies is a measure of the relative depletion and replenishment of molecular gas, and is expected to be a function of halo mass and redshift from simulations¹⁴. In HFLS 3, we find a high gas mass fraction of $f_{\text{gas}} = M_{\text{gas}}/M_{\text{dyn}} \approx 40\%$, comparable to what is found in submillimetre-selected starbursts and massive star-forming galaxies at $z \approx 2$ (refs 15, 16), but ~ 3 times higher than in nearby ultra-luminous infrared galaxies (ULIRGs) like Arp 220, and >30 times higher than in the Milky Way. From population synthesis modelling, we find a stellar mass of $M_* = 3.7 \times 10^{10} M_{\text{sun}}$, comparable to that of Arp 220 and about half that of the Milky Way. This suggests that at most $\sim 40\%$ of M_{dyn} within the radius of the gas reservoir is due to dark matter. With up to $\sim 10^{11} M_{\text{sun}}$ of dark matter within 3.4 kpc, HFLS 3 is likely to reside in a dark-matter halo massive enough to grow a present-day galaxy cluster¹⁷. The efficiency of star formation is given by $\epsilon = t_{\text{dyn}} \times \text{SFR}/M_{\text{gas}}$, where $t_{\text{dyn}} = (r^3/(2GM))^{1/2}$ is the dynamical (or free-fall) time, r is the source radius,

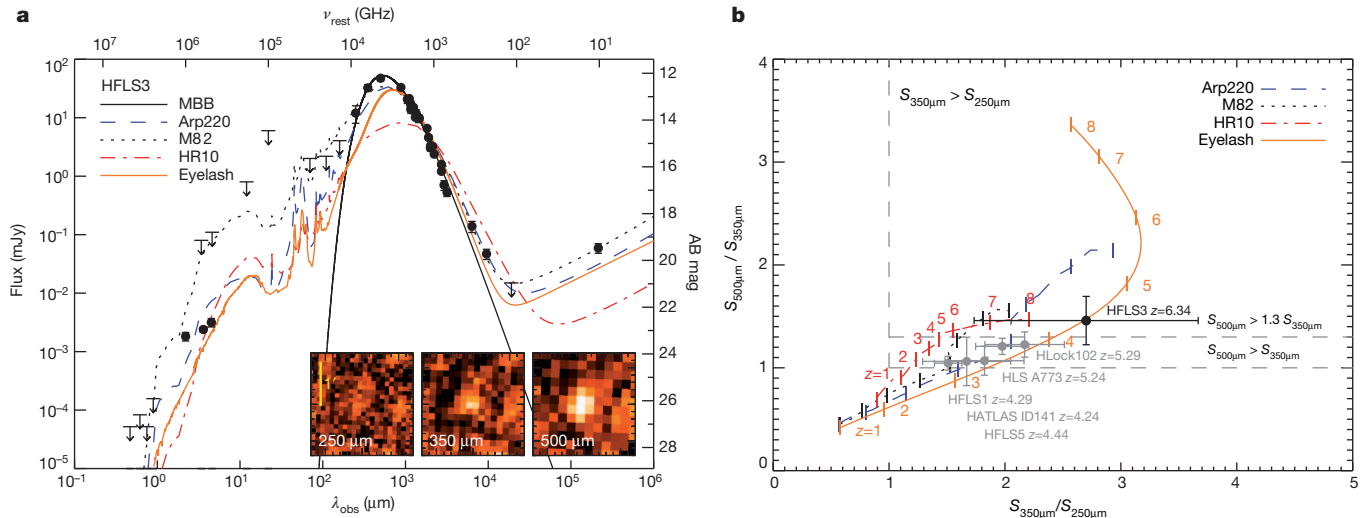


Figure 2 | Spectral energy distribution and Herschel/SPIRE colours of HFLS 3. **a**, HFLS 3 was identified as a very high redshift candidate, as it appears red between the Herschel/SPIRE 250-, 350- and 500- μm bands (inset). The spectral energy distribution of the source (data points; λ_{obs} , observed-frame wavelength; ν_{rest} , rest-frame frequency; AB mag, magnitudes in the AB system; error bars are 1σ r.m.s. uncertainties in both panels) is fitted with a modified black body (MBB; solid line) and spectral templates for the starburst galaxies Arp 220, M 82, HR 10 and the Eyelash (broken lines, see key). The implied FIR luminosity is $2.86^{+0.32}_{-0.31} \times 10^{13} L_{\text{sun}}$. The dust in HFLS 3 is not optically thick at wavelengths longward of rest-frame $162.7 \mu\text{m}$ (95.4% confidence; Supplementary Fig. 12). This is in contrast to Arp 220, in which the dust becomes optically thick (that is, $\tau_d = 1$) shortward of $234 \pm 3 \mu\text{m}$ (ref. 20). Other high-redshift massive starburst galaxies (including the Eyelash) typically become optically thick around $\sim 200 \mu\text{m}$. This suggests that none of the

detected molecular/fine-structure emission lines in HFLS 3 require correction for extinction. The radio continuum luminosity of HFLS 3 is consistent with the radio-FIR correlation for nearby star-forming galaxies. **b**, Flux density ratios ($350 \mu\text{m}/250 \mu\text{m}$ and $500 \mu\text{m}/350 \mu\text{m}$) of HFLS 3. The coloured lines are the same templates as in **a**, but redshifted between $1 < z < 8$ (number labels indicate redshifts). Dashed grey lines indicate the dividing lines for red ($S_{250\mu\text{m}} < S_{350\mu\text{m}} < S_{500\mu\text{m}}$) and ultra-red ($S_{250\mu\text{m}} < S_{350\mu\text{m}}$ and $1.3 \times S_{350\mu\text{m}} < S_{500\mu\text{m}}$) sources. Grey symbols show the positions of five spectroscopically confirmed red sources at $4 < z < 5.5$ (including three new sources from our study), which all fall outside the ultra-red cut-off. This shows that ultra-red sources will lie at $z > 6$ for typical shapes of the spectral energy distribution (except those with low dust temperatures), whereas red sources typically are at $z < 5.5$. See Supplementary Information sections 1 and 3 for more details.

M is the mass within radius r and G is the gravitational constant. For $r = 1.7 \text{ kpc}$ and $M = M_{\text{gas}}$, this suggests $\varepsilon = 0.06$, which is a few times higher than found in nearby starbursts and in giant molecular cloud cores in the Galaxy¹⁸.

The properties of atomic and molecular gas in HFLS 3 are fully consistent with a highly enriched, highly excited interstellar medium,

Table 1 | Observed and derived quantities for HFLS 3, Arp 220 and the Milky Way

	HFLS 3	Arp 220*	Milky Way*
z	6.3369	0.0181	
$M_{\text{gas}} (M_{\text{sun}})^\dagger$	$(1.04 \pm 0.09) \times 10^{11}$	5.2×10^9	2.5×10^9
$M_{\text{dust}} (M_{\text{sun}})^\ddagger$	$1.31^{+0.32}_{-0.30} \times 10^9$	$\sim 1 \times 10^8$	$\sim 6 \times 10^7$
$M_* (M_{\text{sun}})^\S$	$\sim 3.7 \times 10^{10}$	$\sim (3-5) \times 10^{10}$	$\sim 6.4 \times 10^{10}$
$M_{\text{dyn}} (M_{\text{sun}})^\parallel$	2.7×10^{11}	3.45×10^{10}	$2 \times 10^{11} (< 20 \text{ kpc})$
$f_{\text{gas}} (\%)^\P$	40	15	1.2
$L_{\text{FIR}} (L_{\text{sun}})^\#$	$2.86^{+0.32}_{-0.31} \times 10^{13}$	1.8×10^{12}	1.1×10^{10}
SFR ($M_{\text{sun}} \text{ yr}^{-1}$) **	2,900	~ 180	1.3
$T_{\text{dust}} (\text{K})^{**}$	$55.9^{+9.3}_{-12.0}$	66	~ 19

For details see Supplementary Information section 3.

*Literature values for Arp 220 and the Milky Way are adopted from refs 20 and 27–30. The total molecular gas mass of the Milky Way is uncertain by at least a factor of 2. Quoted dust masses and stellar masses are typically uncertain by factors of 2–3 owing to systematics. The dynamical mass for the Milky Way is quoted within the inner 20 kpc to be comparable to the other systems, not probing the outer regions dominated by dark matter. The dust temperature in the Milky Way varies by at least $\pm 5 \text{ K}$ around the quoted value, which is used as a representative value. Both Arp 220 and the Milky Way are known to contain small fractions of significantly warmer dust. All errors are 1σ r.m.s. uncertainties. † Molecular gas mass, derived assuming $x_{\text{CO}} = M_{\text{gas}}/L_{\text{CO}} = 1 M_{\text{sun}} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ (see Supplementary Information section 3.3).

‡ Dust mass, derived from spectral energy distribution fitting (see Supplementary Information section 3.1).

§ Stellar mass, derived from population synthesis fitting (see Supplementary Information section 3.4).

$^\parallel$ Dynamical mass (see Supplementary Information section 3.5).

¶ Gas mass fraction, derived assuming $f_{\text{gas}} = M_{\text{gas}}/M_{\text{dyn}}$ (see Supplementary Information section 3.6).

$^\#$ FIR luminosity as determined over the range of $42.5\text{--}122.5 \mu\text{m}$ from spectral energy distribution fitting (see Supplementary Information section 3.1).

** SFR, derived assuming $\text{SFR} (\text{in } M_{\text{sun}} \text{ yr}^{-1}) = 1.0 \times 10^{-10} L_{\text{FIR}} (\text{in } L_{\text{sun}})$ (see Supplementary Information section 3.2).

** Dust temperature, derived from spectral energy distribution fitting (see Supplementary Information section 3.1).

as typically found in the nuclei of warm, intense starbursts, but distributed over a large, $\sim 3.5\text{-kpc}$ -diameter, region. The observed CO and [C II] luminosities suggest that dust is the primary coolant of the gas if both are thermally coupled. The $L_{[\text{C II}]} / L_{\text{FIR}}$ ratio of $\sim 5 \times 10^{-4}$ is typical for high radiation environments in extreme starbursts and active galactic nucleus (AGN) host galaxies¹⁹. The $L_{[\text{C II}]} / L_{\text{CO}(1-0)}$ ratio of $\sim 3,000$ suggests that the bulk of the line emission is associated with the photon-dominated regions of a massive starburst. At the L_{FIR} of HFLS 3, this suggests an infrared radiation field strength and gas density comparable to nearby ULIRGs without luminous AGN (figures 4 and 5 of ref. 19).

From the spectral energy distribution of HFLS 3, we derive a dust temperature of $T_{\text{dust}} = 56^{+9}_{-12} \text{ K}$, $\sim 10 \text{ K}$ less than in Arp 220, but ~ 3 times that of the Milky Way. CO radiative transfer models assuming collisional excitation suggest a gas kinetic temperature of $T_{\text{kin}} = 144^{+59}_{-30} \text{ K}$ and a gas density of $\log_{10}(n(\text{H}_2)) = 3.80^{+0.28}_{-0.17} \text{ cm}^{-3}$ (Supplementary Information section 4 and Supplementary Figs 13 and 14). These models suggest similar gas densities as in nearby ULIRGs, and prefer $T_{\text{kin}} \gg T_{\text{dust}}$, which may imply that the gas and dust are not in thermal equilibrium, and that the excitation of the molecular lines may be partially supported by the underlying infrared radiation field. This is consistent with the finding that we detect H_2O and OH lines with upper level energies of $E/k_B > 300\text{--}450 \text{ K}$ and critical densities of $> 10^{8.5} \text{ cm}^{-3}$ at line intensities exceeding those of the CO lines. The intensities and ratios of the detected H_2O lines cannot be reproduced by radiative transfer models assuming collisional excitation, but are consistent with being radiatively pumped by FIR photons, at levels comparable to those observed in Arp 220 (Supplementary Figs 15 and 16)^{20,21}. The CO and H_2O excitation is inconsistent with what is observed in quasar host galaxies like Mrk 231 and APM 08279+5255 at $z = 3.9$, which lends support to the conclusion that the gas is excited by a mix of collisions and infrared photons associated with a massive, intense starburst, rather than hard radiation associated with a luminous AGN²². The physical

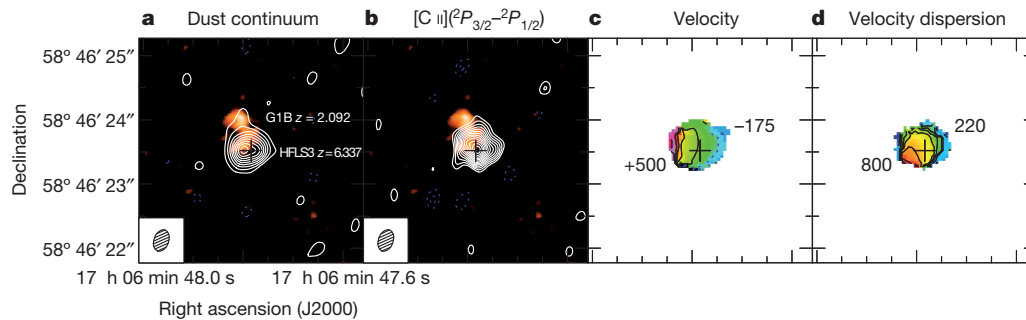


Figure 3 | Gas dynamics, dust obscuration, and distribution of gas and star formation in HFLS 3. **a, b**, High-resolution (FWHM $0.35'' \times 0.23''$) maps of the 158- μm continuum (**a**) and [C II] line emission (**b**) obtained at 1.16 mm with the PdBI in A-configuration, overlaid on a Keck/NIRC2 2.2- μm adaptive optics image (rest-frame ultraviolet/optical light). The r.m.s. uncertainty in the continuum (**a**) and line (**b**) maps is 180 and 400 μJy per beam, and contours are shown in steps of 3σ and 1σ , starting at 5σ and 3σ , respectively. A $z = 2.092$ galaxy (labelled G1B) identified through Keck/LRIS spectroscopy is detected $\sim 0.65''$ north of HFLS 3, but is not massive enough to cause significant gravitational lensing at the position of HFLS 3. Faint infrared emission is detected towards a region with lower dust obscuration in the northeastern part of HFLS 3 (not detected at $< 1 \mu\text{m}$). The Gaussian diameters of the resolved [C II] and continuum emission are $3.4 \text{ kpc} \times 2.9 \text{ kpc}$ and $2.6 \text{ kpc} \times 2.4 \text{ kpc}$,

conditions in the ISM of HFLS 3 thus are comparable to those in the nuclei of the most extreme nearby starbursts, consistent with the finding that it follows the radio–FIR correlation for star-forming galaxies.

HFLS 3 is rapidly assembling its stellar bulge through star formation at surface densities close to the theoretically predicted limit for ‘maximum starbursts’²³. At a rest-frame wavelength of 158 μm , the FIR emission is distributed over a relatively compact area with $2.6 \text{ kpc} \times 2.4 \text{ kpc}$ physical diameter along its major and minor axes respectively (Fig. 3; as determined by elliptical Gaussian fitting). This suggests an extreme SFR surface density of $\Sigma_{\text{SFR}} \approx 600 M_{\text{sun}} \text{ yr}^{-1} \text{ kpc}^{-2}$ over a 1.3-kpc-radius region, and is consistent with near-Eddington-limited star formation if the starburst disk is supported by radiation pressure²⁴. This suggests the presence of a kiloparsec-scale hyperstarburst similar to that found in the $z = 6.42$ quasar J1148+5251 (ref. 25). Such high Σ_{SFR} are also observed in the nuclei of local ULIRGs such as Arp 220, albeit on scales two orders of magnitude smaller. A starburst at such high Σ_{SFR} may produce strong winds. Indeed, the relative strength and broad, asymmetric profile of the $\text{OH } ^2\Pi_{1/2}(3/2-1/2)$ doublet detected in HFLS 3 may indicate a molecular outflow, reminiscent of the OH outflow in Arp 220²¹.

The identification of HFLS 3 alone is still consistent with the model-predicted space density of massive starburst galaxies at $z > 6$ with $S_{500\mu\text{m}} > 30 \text{ mJy}$ of 0.014 deg^{-2} (ref. 7). This corresponds to only 10^{-3} – 10^{-4} times the space density of Lyman-break galaxies at the same redshift, but is comparable to the space density of the most luminous quasars hosting supermassive black holes (that is, a different population of massive galaxies) at such early cosmic times²⁶. The host galaxies around these very distant supermassive black holes are commonly FIR-luminous, but less intensely star-forming, with typically a few times lower L_{FIR} than ultra-red sources²⁵. This highlights the difference between selecting massive $z > 6$ galaxies at the peak of their star formation activity through L_{FIR} , and at the peak of their black-hole activity through luminous AGN. The substantial population of ultra-red sources discovered with Herschel will be an ideal probe of early galaxy evolution and heavy element enrichment within the first billion years of cosmic time. These galaxies are unlikely to dominate the star formation history of the Universe at $z > 6$ (ref. 5), but they trace the highest peaks in SFR at early epochs. A detailed study of this galaxy population will reveal the mass and redshift distribution, number density and likely environments of such objects, which if confirmed

suggesting gas and SFR surface densities of $\Sigma_{\text{gas}} = 1.4 \times 10^4 M_{\text{sun}} \text{ pc}^{-2}$ and $\Sigma_{\text{SFR}} = 600 M_{\text{sun}} \text{ yr}^{-1} \text{ kpc}^{-2}$ ($\sim 0.6 \times 10^{13} L_{\text{sun}} \text{ kpc}^{-2}$). The high Σ_{SFR} is consistent with a maximum starburst at near-Eddington-limited intensity. Given the moderate optical depth of $\tau_d < \sim 1$ at 158 μm , this estimate is somewhat conservative. **c, d**, Peak velocity (**c**) and FWHM velocity dispersion (**d**) maps of the [C II] emission are obtained by Gaussian fitting to the line emission in each spatial point of the map. Velocity contours are shown in steps of 100 km s^{-1} . High-resolution CO $J = 7-6$ and $10-9$ and $\text{H}_2\text{O } 3_{21-3_{12}}$ observations show consistent velocity profiles and velocity structure (Supplementary Figs 5–7). The large velocity dispersion suggests that the gas dynamics in this system are dispersion-dominated. See Supplementary Information sections 3 and 5 for more details.

in larger numbers may present a stern challenge to current models of early cosmic structure formation.

Received 4 October 2012; accepted 27 February 2013.

- Chapman, S. C. *et al.* A median redshift of 2.4 for galaxies bright at submillimetre wavelengths. *Nature* **422**, 695–698 (2003).
- Capak, P. *et al.* A massive protocluster of galaxies at a redshift of $z \approx 5.3$. *Nature* **470**, 233–235 (2011).
- Walter, F. *et al.* The intense starburst HDF 850.1 in a galaxy overdensity at $z \approx 5.2$ in the Hubble Deep Field. *Nature* **486**, 233–236 (2012).
- Vieira, J. D. *et al.* Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature* **495**, 344–347 (2013).
- Bouwens, R. *et al.* A candidate redshift $z \approx 10$ galaxy and rapid changes in that population at an age of 500 Myr. *Nature* **469**, 504–507 (2011).
- Oliver, S. *et al.* The Herschel Multi-tiered Extragalactic Survey: HerMES. *Mon. Not. R. Astron. Soc.* **424**, 1614–1635 (2012).
- B  thermin, M. *et al.* A unified empirical model for infrared galaxy counts based on the observed physical evolution of distant galaxies. *Astrophys. J.* **757**, L23 (2012).
- Karim, A. *et al.* An ALMA survey of submillimetre galaxies in the Extended Chandra Deep Field South: high resolution 870 μm source counts. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/1210.0249> (2012).
- Bouwens, R. J. *et al.* Galaxies at $z \sim 6$: the UV luminosity function and luminosity density from 506 HUDF, HUDF parallel ACS field, and GOODS i-dropouts. *Astrophys. J.* **653**, 53–85 (2006).
- Robertson, B. *et al.* Early star-forming galaxies and the reionization of the Universe. *Nature* **468**, 49–55 (2010).
- Micha  owski, M. J. *et al.* Rapid dust production in submillimetre galaxies at $z > 4$? *Astrophys. J.* **712**, 942–950 (2010).
- Riechers, D. A. *et al.* Extended cold molecular gas reservoirs in $z \sim 3.4$ submillimetre galaxies. *Astrophys. J.* **739**, L31 (2011).
- Walter, F. *et al.* Molecular gas in the host galaxy of a quasar at redshift $z = 6.42$. *Nature* **424**, 406–408 (2003).
- Lagos, C. & Del P. *et al.* On the impact of empirical and theoretical star formation laws on galaxy formation. *Mon. Not. R. Astron. Soc.* **416**, 1566–1584 (2011).
- Tacconi, L. J. *et al.* Submillimetre galaxies at $z \sim 2$: evidence for major mergers and constraints on lifetimes, IMF, and CO– H_2 conversion factor. *Astrophys. J.* **680**, 246–262 (2008).
- Tacconi, L. J. *et al.* High molecular gas fractions in normal massive star-forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).
- Overzier, R. *et al.* ΛCDM predictions for galaxy protoclusters – I. The relation between galaxies, protoclusters and quasars at $z \sim 6$. *Mon. Not. R. Astron. Soc.* **394**, 577–594 (2009).
- Krumholz, M. R. *et al.* A universal, local star formation law in galactic clouds, nearby galaxies, high-redshift disks, and starbursts. *Astrophys. J.* **745**, 69 (2012).
- Stacey, G. J. *et al.* A 158 μm [C II] line survey of galaxies at $z \sim 1$ –2: an indicator of star formation in the early universe. *Astrophys. J.* **724**, 957–974 (2010).
- Rangwala, N. *et al.* Observations of Arp 220 using Herschel-SPIRE: an unprecedented view of the molecular gas in an extreme star formation environment. *Astrophys. J.* **743**, 94 (2011).

21. Gonzalez-Alfonso, E. *et al.* Herschel/PACS spectroscopy of NGC 4418 and Arp 220: H_2O , H_2^{18}O , OH, ^{18}OH , O I, HCN and NH_3 . *Astron. Astrophys.* **541**, A4 (2012).
22. van der Werf, P. *et al.* Water vapor emission reveals a highly obscured, star-forming nuclear region in the QSO host galaxy APM 08279+5255 at $z = 3.9$. *Astrophys. J.* **741**, L38 (2011).
23. Elmegreen, B. G. Galactic bulge formation as a maximum intensity starburst. *Astrophys. J.* **517**, 103–107 (1999).
24. Thompson, T. *et al.* Radiation pressure-supported starburst disks and active galactic nucleus fueling. *Astrophys. J.* **630**, 167–185 (2005).
25. Walter, F. *et al.* A kiloparsec-scale hyper-starburst in a quasar host less than 1 gigayear after the Big Bang. *Nature* **457**, 699–701 (2009).
26. Jiang, L. *et al.* A survey of $z \sim 6$ quasars in the Sloan Digital Sky Survey deep stripe. II. Discovery of six quasars at $z_{\text{AB}} > 21$. *Astron. J.* **138**, 305–311 (2009).
27. Downes, D. & Solomon, P. M. Rotating nuclear rings and extreme starbursts in ultraluminous galaxies. *Astrophys. J.* **507**, 615–654 (1998).
28. Sodroski, T. J. *et al.* Large-scale characteristics of interstellar dust from COBE DIRBE observations. *Astrophys. J.* **428**, 638–646 (1994).
29. Murray, N. & Rahman, M. Star formation in massive clusters via the Wilkinson Microwave Anisotropy Probe and the Spitzer Glimpse survey. *Astrophys. J.* **709**, 424–435 (2010).
30. McMillan, P. J. Mass models of the Milky Way. *Mon. Not. R. Astron. Soc.* **414**, 2446–2457 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. This research has made use of data from the HerMES project. HerMES is a Herschel Key Programme using guaranteed time from the SPIRE instrument team, ESAC scientists and a mission scientist. See Supplementary Information for further acknowledgements.

Author Contributions D.A.R. had the overall lead of the project. C.M.B., D.L.C., I.P.-F., R.J.I., C.B., H.F., J.D.V. and R.N. contributed significantly to the taking and analysis of the follow-up data with different instruments by leading several telescope proposals and analysis efforts. C.D.D. led the selection of the parent sample. A. Conley, J.W., J.C., A. Cooray, P.H. and J.K. contributed significantly to the data analysis and to fitting and modelling the results. All other authors contributed to the proposals, source selection, data analysis and interpretation, in particular through work on the primary Herschel SPIRE data in which the source was discovered through the HerMES consortium (led by J.B. and S.J.O.). All authors have reviewed, discussed, and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.R. (dr@astro.cornell.edu).

High-fidelity readout and control of a nuclear spin qubit in silicon

Jarryd J. Pla¹, Kuan Y. Tan^{1†}, Juan P. Dehollain¹, Wee H. Lim^{1†}, John J. L. Morton², Floris A. Zwanenburg^{1†}, David N. Jamieson³, Andrew S. Dzurak¹ & Andrea Morello¹

Detection of nuclear spin precession is critical for a wide range of scientific techniques that have applications in diverse fields including analytical chemistry, materials science, medicine and biology. Fundamentally, it is possible because of the extreme isolation of nuclear spins from their environment. This isolation also makes single nuclear spins desirable for quantum-information processing, as shown by pioneering studies on nitrogen-vacancy centres in diamond^{1–4}. The nuclear spin of a ³¹P donor in silicon is very promising as a quantum bit⁵: bulk measurements indicate that it has excellent coherence times^{6,7} and silicon is the dominant material in the microelectronics industry. Here we demonstrate electrical detection and coherent manipulation of a single ³¹P nuclear spin qubit with sufficiently high fidelities for fault-tolerant quantum computing⁸. By integrating single-shot readout of the electron spin⁹ with on-chip electron spin resonance¹⁰, we demonstrate quantum non-demolition¹¹ and electrical single-shot readout of the nuclear spin with a readout fidelity higher than 99.8 per cent—the highest so far reported for any solid-state qubit. The single nuclear spin is then operated as a qubit by applying coherent radio-frequency pulses. For an ionized ³¹P donor, we find a nuclear spin coherence time of 60 milliseconds and a one-qubit gate control fidelity exceeding 98 per cent. These results demonstrate that the dominant technology of modern electronics can be adapted to host a complete electrical measurement and control platform for nuclear-spin-based quantum-information processing.

Quantum computers have the potential to revolutionize aspects of modern society, ranging from fundamental science to information technology¹². The successful demonstration of such a machine requires the ability to perform high-fidelity control and measurement of individual qubits—the building blocks of a quantum computer. Errors introduced by quantum operations and measurements can be mitigated by using quantum error correction protocols, provided that the probabilities of the errors occurring are below certain stringent thresholds⁸. State-of-the-art high-fidelity qubit control and readout has been realized using laser-cooled atoms in electromagnetic traps^{13,14}—a result made possible because of their extreme isolation in a near-perfect vacuum.

Qubits based on solid-state physical systems^{1,15,16} are attractive because of the potential for scalability using modern integrated-circuit fabrication technologies. However, they tend to exhibit much lower system fidelities owing to interactions with their host environment. The ability to combine the control and measurement fidelities of trapped atoms with the scalability inherent to solid-state implementations is therefore highly desirable. The nuclear spin of a single atom is promising in this regard as it is a simple, well-isolated quantum system. One of the earliest proposals for quantum computing in the solid state advocated using the nuclear spin of individual ³¹P dopant atoms in silicon to encode and process quantum information⁵. Silicon is an

excellent platform for spin-based quantum-information processing because it can be enriched in the nuclear-spin-zero ²⁸Si isotope¹⁷, providing an effective ‘semiconductor vacuum’ and very long spin coherence times. Experiments in bulk phosphorus-doped isotopically enriched silicon (²⁸Si:P) have demonstrated the potential of this system in which the ³¹P nuclear spin has been used as a quantum memory⁷ and as a qubit with extraordinarily long coherence lifetimes of over 180 s (ref. 6). However, owing to detection limitations, experiments have so far been carried out on only large ensembles of ³¹P nuclei typically consisting of several billion nuclear spins^{6,18}. To realize nuclear-spin-based solid-state quantum computing, one must first isolate, measure and control individual nuclear spins.

It is challenging to measure the state of a single nuclear spin owing to its weak magnetic moment. In the solid state, this has been achieved only for the nitrogen-vacancy centre in diamond², with optical detection, and for a rare-earth terbium ion by performing electrical transport measurements through a single molecule¹⁹. In both these cases, a coupled electron spin was used to read out the nuclear spin, which requires the ability to measure a single electron spin. The same applies to the ³¹P nucleus: the high-fidelity single-shot readout of the donor-bound electron spin⁹ is at the heart of our nuclear spin readout method.

We use an on-chip all-electrical detection method for both electron and nuclear spins, using a compact nanoscale device²⁰ (Fig. 1a) compatible with silicon metal–oxide–semiconductor fabrication standards. The device consists of ion-implanted phosphorus donors²¹, tunnel coupled to a silicon metal–oxide–semiconductor single-electron transistor (SET)²².

The ³¹P donor in silicon possesses a nuclear spin I of 1/2, and at cryogenic temperatures it can bind an electron (with spin $S = 1/2$) in the neutral D⁰ donor charge state. Therefore, a single ³¹P donor constitutes a two-qubit system, where the two qubits interact with an external magnetic field B_0 in proportion to their gyromagnetic ratios: $\gamma_n = 17.23 \text{ MHz T}^{-1}$ for the nucleus²³ and $\gamma_e = g\mu_B/\hbar = 27.97 \text{ GHz T}^{-1}$ for the electron, where $g = 1.9985$ (ref. 24) is the Landé g -factor, μ_B is the Bohr magneton and \hbar is Planck’s constant. In addition, they interact with each other through the hyperfine interaction $A = 117.53 \text{ MHz}$ (ref. 24) that arises from the overlap between the wavefunctions of the electron and the ³¹P nucleus. If $\gamma_e B_0 \gg A > 2\gamma_n B_0$, the eigenstates of the two-spin system are approximately (in ascending order of energy) $|\uparrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$, $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$, where the thin (thick) arrow indicates the orientation of the electron (nuclear) spin (Fig. 1b). The system can be transformed to that of a single ³¹P nuclear spin (Fig. 1c) with eigenstates $|\uparrow\rangle$ and $|\downarrow\rangle$ by using the nanostructure device (Fig. 1a) to ionize the donor to the D⁺ charge state.

The ³¹P nuclear spin readout experiment begins by performing electron spin resonance (ESR) on its bound donor electron¹⁰, using microwave pulses delivered by an on-chip broadband planar transmission line²⁵. The system exhibits two possible ESR frequencies,

¹Centre for Quantum Computation and Communication Technology, School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, New South Wales 2052, Australia.

²London Centre for Nanotechnology, University College London, London WC1H 0AH, UK. ³Centre for Quantum Computation and Communication Technology, School of Physics, University of Melbourne, Melbourne, Victoria 3010, Australia. [†]Present addresses: QCD Labs, COMP Centre of Excellence, Department of Applied Physics, Aalto University, PO Box 13500, FI-00076 Aalto, Finland (K.Y.T.); Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia (W.H.L.); NanoElectronics Group, MESA+ Institute for Nanotechnology, University of Twente, 7500 AE Enschede, The Netherlands (F.A.Z.).

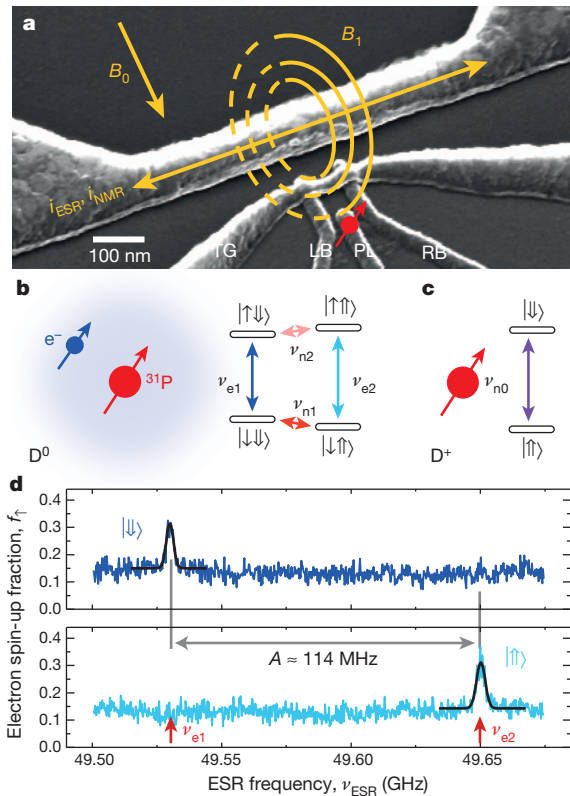


Figure 1 | Qubit nanostructure, spin transitions and electron spin resonance spectra. **a**, Scanning electron micrograph of the active area of the qubit device, showing an implanted donor (donor as red arrow), the single-electron transistor (SET) and the short-circuit termination of the microwave line. The device is mounted in a dilution refrigerator with an electron temperature of ~ 300 mK, and is subjected to static magnetic fields B_0 between 1.0 T and 1.8 T. B_0 is oriented perpendicular to the short-circuit termination of the microwave line (solid orange single-ended arrow), which carries a current (solid double-ended arrow) and produces an oscillating magnetic field B_1 (represented by the solid and dashed circles) perpendicular to the surface of the device. TG, top gate; PL, plunger gate; LB, left barrier; RB, right barrier. **b**, Energy-level diagram of the neutral ^{31}P donor system, with corresponding transitions for electron spin resonance (ESR) in blue, and for nuclear magnetic resonance (NMR) in red. $\uparrow\uparrow$: electron spin states; $\downarrow\downarrow$: nuclear spin states. **c**, Energy-level diagram of the ionized ^{31}P donor, with the single NMR transition shown in purple. **d**, ESR spectra obtained at $B_0 = 1.77$ T by scanning the microwave frequency and monitoring the electron-spin-up fraction f_1 . The top trace corresponds to an active ν_{e1} ESR transition (nuclear spin state $|\downarrow\downarrow\rangle$) and the bottom trace to an active ν_{e2} ESR transition (nuclear spin state $|\uparrow\uparrow\rangle$).

depending on the state of the nuclear spin: $\nu_{e1} \approx \gamma_e B_0 - A/2$ for nuclear spin $|\downarrow\downarrow\rangle$ and $\nu_{e2} \approx \gamma_e B_0 + A/2$ for nuclear spin $|\uparrow\uparrow\rangle$. In a single-atom experiment, if we assume that the ESR measurement duration is much shorter than the nuclear spin flip time, then we expect only one active ESR frequency at any instant. Detecting ESR at the frequency ν_{e1} therefore indicates that the nuclear spin is in state $|\downarrow\downarrow\rangle$, whereas detection at ν_{e2} implies the nuclear spin is $|\uparrow\uparrow\rangle$.

Having identified the two resonance frequencies through an ESR experiment (see Fig. 1d and also ref. 10), we performed repeated measurements of the nuclear spin state (Fig. 2a) by toggling the microwave frequency ν_{ESR} between ν_{e1} and ν_{e2} , obtaining the electron-spin-up fraction f_1 at each point (see Supplementary Information). If the quantity $\Delta f_1 = f_1(\nu_{e2}) - f_1(\nu_{e1})$ is positive, we assign the nuclear state $|\uparrow\uparrow\rangle$, and vice versa. A histogram of Δf_1 (Fig. 2d) shows two well-separated Gaussian peaks, corresponding to the two possible nuclear orientations, with widths determined by the signal-to-noise ratio (SNR) of the measurements (Supplementary Information). The nuclear spin readout error (Fig. 2e) is obtained by fitting the two peaks and integrating

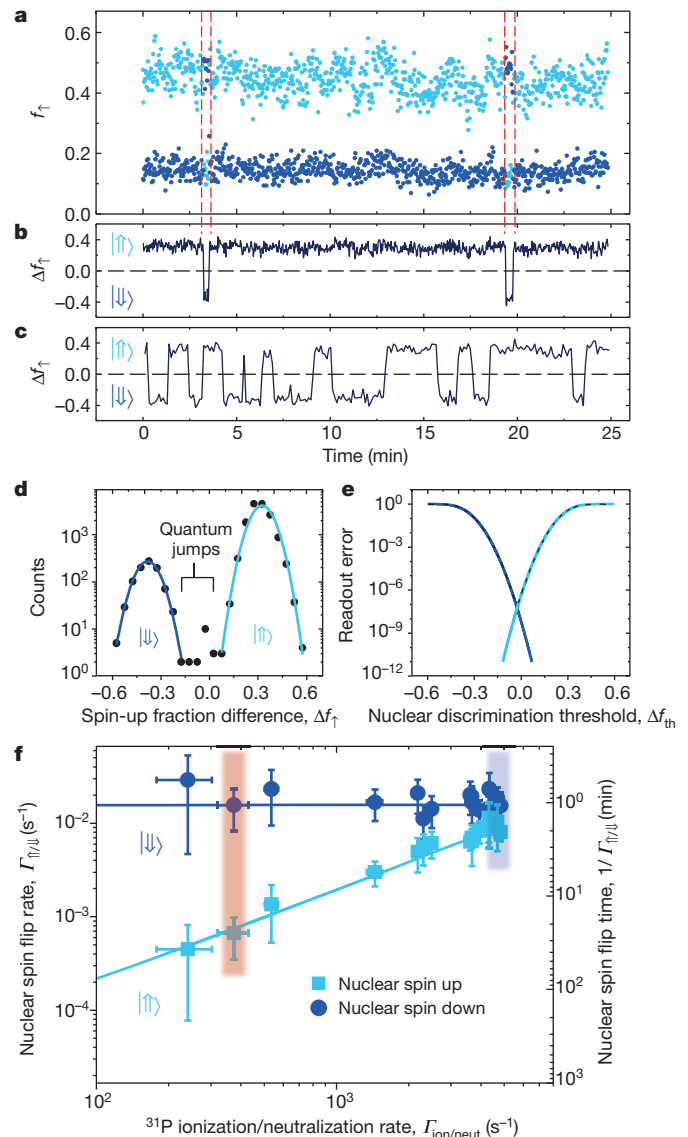


Figure 2 | Nuclear spin quantum jumps, readout error and lifetimes.

a, Repetitive single-shot measurements of the nuclear spin state performed by toggling ν_{ESR} between $\nu_{e1} = 49.5305$ GHz (dark blue) and $\nu_{e2} = 49.6445$ GHz (light blue) and recording the electron-spin-up fraction f_1 . Each data point represents the average f_1 over 250 single-shot measurements of the electron spin (acquired in 260 ms). The electron-tunnelling time is of the order of 100 μs , as reported in ref. 10. The dashed red lines indicate instants when a nuclear spin quantum jump has occurred. **b**, Electron-spin-up fraction difference, $\Delta f_1 = f_1(\nu_{e2}) - f_1(\nu_{e1})$, for the data in **a**. $\Delta f_1 > 0$ indicates nuclear spin $|\uparrow\uparrow\rangle$, and vice versa. **c**, Δf_1 in an experiment with an increased rate of donor ionization/neutralization, $\Gamma_{\text{ion/neut}}$. The greater $\Gamma_{\text{ion/neut}}$ is achieved by including an additional phase in the nuclear spin readout measurement, where resonant tunnelling of $|\downarrow\downarrow\rangle$ electrons between the donor and SET occurs (see Supplementary Information). **d**, Histograms of Δf_1 for the data in **b**, showing two well-separated Gaussian peaks, each corresponding to a nuclear spin state, as indicated. The counts obtained for $-0.015 < \Delta f_1 < 0.05$ are attributed to nuclear spin quantum jumps occurring during the measurement. The light and dark blue solid lines are Gaussian fits to the data (see discussion in main text). **e**, Readout errors as a function of the detection threshold for Δf_1 . The solid dark (light) blue line indicates the SNR-limited error for detecting the $|\downarrow\downarrow\rangle$ ($|\uparrow\uparrow\rangle$) state, whereas the black dashed line indicates the total error. **f**, Nuclear spin flip rates $\Gamma_{\uparrow/\downarrow}$ as a function of the donor ionization/neutralization rate $\Gamma_{\text{ion/neut}}$. The light blue line is a fit to $\Gamma_{\uparrow} = \Gamma_0 + p\Gamma_{\text{ion/neut}}$, with $p = 1.91(8) \times 10^{-6}$. The dark blue line is a constant $\Gamma_{\downarrow} = 1.54(17) \times 10^{-2} \text{ s}^{-1}$. The red and blue shaded regions indicate the values obtained from the data sets in **b** and **c**, respectively. The error bars represent a 95% confidence level (see Supplementary Information).

each Gaussian beyond a discrimination threshold Δf_{th} . At the optimal value of $\Delta f_{\text{th}} = -0.025$, the SNR-limited readout error is 2×10^{-7} .

We observe that the nuclear spin state remains unchanged for several minutes before exhibiting a ‘quantum jump’ to the opposite state². It is also evident (Fig. 2b) that the nuclear spin is predominantly oriented in the $|\uparrow\rangle$ state, which we attribute to an electron–nuclear spin flip-flop process, in which the energy difference $E_{\uparrow\downarrow} - E_{\downarrow\uparrow}$ (that is, between states $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$) is released to the phonon bath^{18,26}. Because $E_{\uparrow\downarrow} - E_{\downarrow\uparrow} \gg k_B T$ in our experiment, this process acts only in the direction $|\uparrow\downarrow\rangle \rightarrow |\downarrow\uparrow\rangle$ (that is, only spontaneous emission of phonons occurs), and it cannot be responsible for the observed nuclear spin jumps from $|\uparrow\rangle$ to $|\downarrow\rangle$. We have verified that the $|\uparrow\rangle \rightarrow |\downarrow\rangle$ transition originates from the readout process, where the donor undergoes repeated ionization and neutralization events. These events result in a

time-varying Hamiltonian, where the exact nature of the nuclear eigenstates varies slightly between the neutral and the ionized donor case. Accordingly, we observe (Fig. 2c, f) that the lifetime of the nuclear spin $|\uparrow\rangle$ is inversely proportional to the rate of ionization and neutralization (see also Supplementary Information).

By exploiting the broadband nature of our on-chip microwave transmission line, we perform a nuclear magnetic resonance (NMR) experiment (see Supplementary Information for details) on the ^{31}P nuclear spin (Fig. 3a). For a neutral donor, we expect two NMR frequencies depending on the state of the electron: $\nu_{n1} = A/2 + \gamma_n B_0$ when the electron spin is $|\downarrow\rangle$ and $\nu_{n2} = A/2 - \gamma_n B_0$ when the electron spin is $|\uparrow\rangle$ (Fig. 1b). However, we can also perform an NMR experiment while the donor is ionized (Fig. 1c), as recently demonstrated in a bulk Si:P sample²⁷. In this case there is only one resonance frequency:

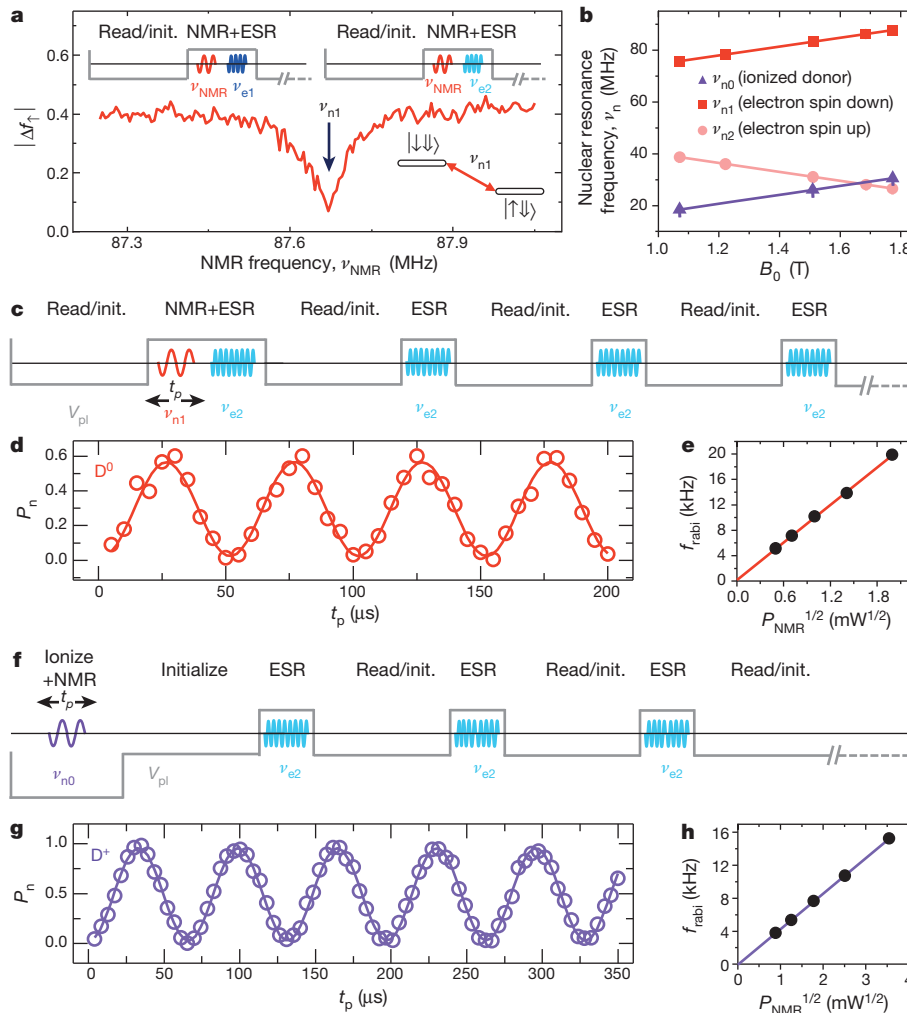


Figure 3 | NMR spectra and Rabi oscillations of a single ^{31}P nuclear spin.

a, Observation of nuclear resonance at $B_0 = 1.77$ T, while the electron spin is initialized $|\downarrow\rangle$ (see Supplementary Information). The top inset shows the plunger gate voltage waveform (grey line) plus NMR/ESR pulse sequence, and the bottom-right inset shows the energy levels involved in the NMR transition. **b**, Dependence of the NMR resonances ν_{n0} , ν_{n1} and ν_{n2} on the magnetic field B_0 . Solid lines are the values predicted using the ^{31}P nuclear gyromagnetic ratio $\gamma_n = 17.23 \text{ MHz T}^{-1}$. **c**, Pulse sequence for the coherent rotation of a ^{31}P nuclear spin with the donor in the neutral D^0 state. Depicted is the plunger gate voltage waveform (grey line) and combined NMR/ESR pulses. After a coherent NMR pulse at ν_{n1} of duration t_p , the nuclear spin state is read by probing the ν_{e2} ESR transition with 300 single-shot adiabatic inversion and electron-spin readout measurements, lasting approximately 300 ms. The resulting electron-spin-up fraction $f_e(\nu_{e2})$ is compared to a threshold, extracted from the quantum jump experiment (Fig. 2a), and a nuclear spin orientation is ascribed to the measurement. **d**, Rabi oscillation

of the neutral ^{31}P donor nuclear spin with $P_{\text{NMR}} = 4 \text{ mW}$, $B_0 = 1.07 \text{ T}$ and $\nu_{n1} = 75.7261 \text{ MHz}$. The pulse sequence of **c** is repeated 40 times for each Rabi pulse length t_p , with five sweeps of t_p performed to give a total of 200 measurements at each t_p . The number of nuclear spin flips is recorded to give the flip probability P_n . The solid line is a fit of the form $P_n = K \sin^2(\pi f_{\text{Rabi}} t_p)$, where K and f_{Rabi} are free fitting parameters. **e**, Rabi frequency, extracted from fits of data similar to that in **d**, against the square root of the radio-frequency power. **f**, Modified pulse sequence to perform Rabi oscillations on the ^{31}P nuclear spin with the donor in the ionized D^+ state. The electron is removed before applying a coherent NMR burst. The electron is then replaced so that a single-shot measurement can be performed on the nuclear spin. **g**, Sample Rabi oscillation of the ionized donor nuclear spin using $P_{\text{NMR}} = 126 \text{ mW}$, $B_0 = 1.77 \text{ T}$ and $\nu_{n0} = 30.5485 \text{ MHz}$, with each data point again obtained from 200 measurements of the nuclear spin state. **h**, Plot showing the linear variation of the ionized nuclear spin Rabi frequency with the excitation amplitude.

$\nu_{n0} = \gamma_n B_0$. Figure 3b shows the magnetic-field dependence of the three NMR frequencies, which agree with the expected values assuming the bulk ^{31}P gyromagnetic ratio $\gamma_n = 17.23 \text{ MHz T}^{-1}$ (ref. 23). Furthermore, we extract $g = 1.9987(6)$ (see Supplementary Information), which is within about 0.01% of the bulk value for Si:P, whereas the hyperfine splitting $A = 114.30(1) \text{ MHz}$ is Stark shifted⁵ from the bulk value of 117.53 MHz (ref. 24). The observation of a Stark shift of A is important, as it has been proposed⁵ as a mechanism to address individual ^{31}P nuclear spin qubits while applying a global microwave field.

Having found the NMR frequencies, we use short radio-frequency pulses to produce coherent superpositions of the nuclear spin states. For the neutral (D^0) donor, we first initialize the electron in the $|\downarrow\rangle$ state. A pulse of length t_p and at the ν_{n1} resonance is applied immediately after, followed by a single-shot readout of the nuclear spin state (see Fig. 3c). Measuring the nuclear spin flip probability P_n as a function of t_p results in the coherent Rabi oscillations of Fig. 3d, whose frequency f_{rabi} scales linearly with the square root of the radio-frequency excitation power, $P_{\text{NMR}}^{1/2}$ (Fig. 3e). The visibility of the oscillations in Fig. 3d is $\sim 60\%$. Deviations from ideality are most probably due to erroneous electron initialization in the excited $|\uparrow\rangle$ state¹⁰, caused by heating resulting from the train of microwave pulses used during readout. We modified the pulse sequence to remove the electron before applying the radio-frequency excitation at the ν_{n0} transition (Fig. 3f). The Rabi oscillations on the ionized (D^+) ^{31}P nuclear spin (Fig. 3g, h) have near-unity visibility, as the electron spin state has no bearing on the nuclear resonance frequency while the donor is ionized.

To assess the viability of using the ^{31}P nuclear spin as a quantum bit, it is critical to characterize the duration over which coherence is preserved. The dephasing time T_2^* is obtained from a Ramsey-fringe measurement (see Supplementary Information), the NMR pulse sequence for which is shown in Fig. 4a (see also Fig. 4b for a Bloch sphere state evolution). The decay of the oscillations in Fig. 4c, with increasing inter-pulse delay τ , is the result of fluctuations in the local magnetic environment. Fitting the data with a damped cosine function $P_n(\tau) = P_n(0)\cos(2\pi\Delta\tau)\exp(-\tau/T_2^*)$, where $P_n(0)$ is the amplitude and $\Delta\tau$ is the average detuning from resonance, reveals $T_2^*(\text{D}^0) = 0.84(10) \text{ ms}$ for the neutral donor and $T_2^*(\text{D}^+) = 3.3(3) \text{ ms}$ for the ionized donor. These dephasing times are $\sim 10^4$ times longer than those measured for the electron spin¹⁰.

As many of the magnetic fluctuations that contribute to T_2^* occur on much longer timescales than the typical nuclear spin manipulation time ($\sim 25 \mu\text{s}$ for a π pulse), a significant portion of the dephasing can be reversed by using spin-echo techniques (Fig. 4d, e). Observing the echo signal while varying the delay τ yields the decay curves displayed in Fig. 4f. We fit the data with functions of the form $y = y(0)\exp[(-2\tau/T_2)^b]$, where $y(0)$ is the amplitude, b is a free exponent and T_2 is the coherence time. For the ionized donor spin, we extract $T_2(\text{D}^+) = 60.0(9) \text{ ms}$ and $b(\text{D}^+) = 1.77(7)$, both of which are fully accounted for by the spectral diffusion caused by dynamics of the ^{29}Si nuclear spin bath, as quantified by recent theory²⁸. Accordingly, we expect that removal of ^{29}Si through isotopic purification¹⁷ should yield T_2 values of the order of minutes, as observed in bulk-doped samples⁶.

Conversely, for the neutral donor spin, we find a shorter $T_2(\text{D}^0) = 3.5(1) \text{ ms}$ and $b(\text{D}^0) = 2.2(2)$. This suggests that additional decoherence processes occur in the presence of the donor-bound electron. Charge noise at the Si/SiO₂ interface²⁷ or electronic gate noise⁵ could cause a time-dependent Stark shift of the hyperfine coupling, resulting in fluctuations of the instantaneous nuclear Larmor frequency. Ionizing the donor thus removes the connection between the electrical noise and the nuclear spin precession frequency. This is a static effect, in contrast to a recent experiment on the nitrogen-vacancy centre in diamond³, where rapid ionization and neutralization improved the coherence of a ^{13}C nuclear spin through motional averaging.

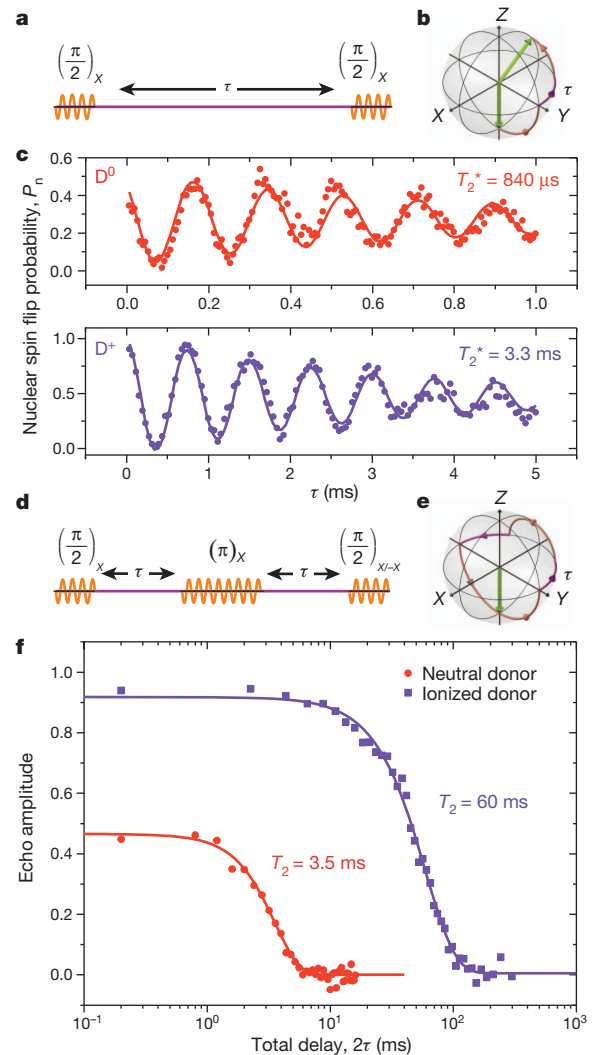


Figure 4 | Ramsey fringes and spin-echo decay. **a**, NMR pulse sequence for the Ramsey-fringe experiment. This sequence replaces the single pulse of duration t_p in Fig. 3c (Fig. 3f) for the neutral (ionized) donor, whereas the nuclear spin is read out in the same way. The phase of both $\pi/2$ pulses is such that rotation is performed about the X axis on the Bloch sphere, as noted in the rotation angle subscript above each pulse. **b**, Bloch-sphere representation of the evolution in the rotating frame for the Ramsey-fringe measurement. The green arrow represents the nuclear spin. The purple path indicates dephasing in between pulses, whereas the orange path represents a rotation about the X axis. **c**, Ramsey interference fringes for the nuclear spin with the donor in the D^0 (top) and D^+ (bottom) charge states, taken at $B_0 = 1.77 \text{ T}$. Here a $\pi/2$ pulse was $12.5 \mu\text{s}$ for the D^0 experiment and $23.5 \mu\text{s}$ for the D^+ experiment. We sweep the inter-pulse delay, and repeat the sequence 20 times at each τ . Ten sweeps are performed (with the total 200 measurements occurring over $\sim 3 \text{ min}$ for each τ) and the nuclear spin flip probability P_n is found. Fits to the data are discussed in the main text. **d**, NMR pulse sequence for the spin-echo experiment. Here we also implement phase cycling, where the final $\pi/2$ rotation is first performed about the X axis and the measurement is then repeated with the final $\pi/2$ rotation about the $-X$ axis. Subtracting the two signals ensures a baseline of zero. **e**, Bloch-sphere representation of the spin-echo measurement. Here the final $\pi/2$ pulse is about the X axis (the $-X$ axis is not shown). **f**, Decay of the echo amplitude as the delay τ is increased for the case of a neutral (circles) and ionized (squares) donor. We perform 40 repetitions of the sequence for each τ and 25 sweeps, totalling 1,000 measurements at each point. The phase-cycled echo amplitude is given by $P_n(\nu_n, -X) - P_n(\nu_n, X)$, where $P_n(\nu_n, -X/X)$ represents the nuclear spin flip probability measured at the NMR resonance ν_n with a final $\pi/2$ pulse about the $-X$ or X axis. All other experimental conditions are as in the Ramsey-fringe experiment. Fits through the data are discussed in the text.

We now analyse the fidelity of our solid-state qubit. Single-shot nuclear spin readout can be performed with extremely high fidelity, owing to the quantum non-demolition (QND)¹¹ nature of the measurement. The physical phenomena responsible for the observed quantum jumps of the nuclear spin originate from the measurement through the electron spin, and can be viewed as a deviation from QND ideality (see Supplementary Information). For the nuclear $|\uparrow\rangle$ state, this results in a lifetime of $T_{\uparrow} = 1,500(360)$ s (obtained from extended data of the measurement in Fig. 2a). For the nuclear $|\downarrow\rangle$ state, the cross-relaxation process, which is caused by phonons modulating the hyperfine coupling, yields $T_{\downarrow} = 65(15)$ s (Fig. 2a). These lifetimes must be contrasted with the nuclear-spin measurement time T_{meas} , which has been optimized here to maximize the nuclear-spin readout fidelity (see Supplementary Information). Combining the optimal measurement time ($T_{\text{meas}} = 104$ ms) with the observed nuclear-spin lifetimes yields the QND fidelities: $F_{\text{QND}}(|\uparrow\rangle) = \exp(-T_{\text{meas}}/T_{\uparrow}) = 0.99993(2)$; and $F_{\text{QND}}(|\downarrow\rangle) = \exp(-T_{\text{meas}}/T_{\downarrow}) = 0.9984(4)$. We have therefore obtained readout fidelities between 99.8% and 99.99%, the highest reported so far for any solid-state qubit, and comparable with the fidelities observed for qubits in vacuum-based ion-trap systems¹³.

The nuclear spin control fidelity F_C will ultimately be limited by T_2 and the minimum achievable gate-operation time. We quantify F_C by directly measuring the rotation-angle error for the ionized donor nuclear spin using multiple-pulse dynamical-decoupling sequences (see Supplementary Information). Such a measurement encompasses extrinsic sources of pulse error, for example due to slow power fluctuations of the radio-frequency source. The extracted maximum uncertainty of 3° for an intended π pulse indicates a lower bound on F_C of 98%.

The results presented here demonstrate that the nuclear spin of a single ^{31}P donor could constitute a quantum memory in an electron-spin-based quantum computer¹⁰, or a high-fidelity qubit in a quantum processor where the nuclear spin is the primary computational element⁵. Future experiments will therefore focus on demonstrating electron–nuclear entanglement^{29,30} and the coupling of multiple nuclei mediated by hyperfine and exchange interactions⁵. We anticipate that exploiting the ^{31}P nuclear spin qubit will open new avenues for large-scale quantum-computer architectures, where the quantum coherence of well-isolated atomic systems is combined with the manufacturability of silicon nanoelectronic devices.

METHODS SUMMARY

Device fabrication. For information relating to the device fabrication, see ref. 10, where it is described in some detail.

Experimental set-up. For our voltage pulses, we employed a compensation technique—using an arbitrary waveform generator (Tektronix, AWG520) and an inverting voltage preamplifier—to ensure that the pulsing shifted only the donor electrochemical potentials while keeping the SET island potential constant. The voltage V_p (see Supplementary Fig. 1b) was applied directly to the top gate, whereas it was inverted and amplified by a factor K before reaching the plunger gate. The gain K was carefully tuned to ensure that the SET operating point moved along the top of the SET current peaks, as indicated by the blue arrow in Fig. 1d of ref. 9. The SET current was measured by a Femto DLPCA-200 transimpedance amplifier at room temperature, followed by a voltage post-amplifier, a sixth-order low-pass Bessel filter and a fast digitizing oscilloscope.

The ESR excitations were produced by an Agilent E8257D microwave analogue signal generator and the NMR excitations by an Agilent MXG N5182A radio-frequency vector signal generator. The two signals were combined at room temperature using a power divider/combiner, before being guided to the sample by a semi-rigid coaxial cable (2.2 m in length). Gating of the ESR/NMR pulses was provided by the Tektronix AWG520, which was synchronized with the top gate and plunger gate pulses.

Received 9 December 2012; accepted 13 February 2013.

1. Jelezko, F. *et al.* Observation of coherent oscillation of a single nuclear spin and realization of a two-qubit conditional quantum gate. *Phys. Rev. Lett.* **93**, 130501 (2004).

2. Neumann, P. *et al.* Single-shot readout of a single nuclear spin. *Science* **329**, 542–544 (2010).
3. Maurer, P. C. *et al.* Room-temperature quantum bit memory exceeding one second. *Science* **336**, 1283–1286 (2012).
4. Pfaff, W. *et al.* Demonstration of entanglement-by-measurement of solid-state qubits. *Nature Phys.* **9**, 29–33 (2013).
5. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
6. Steger, M. *et al.* Quantum information storage for over 180 s using donor spins in a ^{28}Si 'semiconductor vacuum'. *Science* **336**, 1280–1283 (2012).
7. Morton, J. J. L. *et al.* Solid-state quantum memory using the ^{31}P nuclear spin. *Nature* **455**, 1085–1088 (2008).
8. Knill, E. Quantum computing with realistically noisy devices. *Nature* **434**, 39–44 (2005).
9. Morello, A. *et al.* Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
10. Pla, J. J. *et al.* A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).
11. Braginsky, V. B., Khalili, F. & Ya. Quantum nondemolition measurements: the route from toys to tools. *Rev. Mod. Phys.* **68**, 1–11 (1996).
12. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
13. Myerson, A. H. *et al.* High-fidelity read-out of trapped-ion qubits. *Phys. Rev. Lett.* **100**, 200502 (2008).
14. Brown, K. R. *et al.* Single-qubit-gate error below 10^{-4} in a trapped ion. *Phys. Rev. A* **84**, 030303(R) (2011).
15. Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
16. Hanson, R. & Awschalom, D. D. Coherent manipulation of single spins in semiconductors. *Nature* **453**, 1043–1049 (2008).
17. Becker, P., Pohl, H. J., Riemann, H. & Abrosimov, N. Enrichment of silicon for a better kilogram. *Phys. Status Solidi A* **207**, 49–66 (2010).
18. McCarney, D. R., Van Tol, J., Morley, G. W. & Boehme, C. Electronic spin storage in an electrically readable nuclear spin memory with a lifetime > 100 seconds. *Science* **330**, 1652–1656 (2010).
19. Vincent, R., Klyatskaya, S., Ruben, M., Wernsdorfer, W. & Balestro, F. Electronic read-out of a single nuclear spin using a molecular spin transistor. *Nature* **488**, 357–360 (2012).
20. Morello, A. *et al.* Architecture for high-sensitivity single-shot readout and control of the electron spin of individual donors in silicon. *Phys. Rev. B* **80**, 081307(R) (2009).
21. Jamieson, D. N. *et al.* Controlled shallow single-ion implantation in silicon using an active substrate for sub-20-keV ions. *Appl. Phys. Lett.* **86**, 202101 (2005).
22. Angus, S. J., Ferguson, A. J., Dzurak, A. S. & Clark, R. G. Gate-defined quantum dots in intrinsic silicon. *Nano Lett.* **7**, 2051–2055 (2007).
23. Steger, M. *et al.* Optically-detected NMR of optically-hyperpolarized ^{31}P neutral donors in ^{28}Si . *J. Appl. Phys.* **109**, 102411 (2011).
24. Feher, G. Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. *Phys. Rev.* **114**, 1219–1244 (1959).
25. Dehollain, J. P. *et al.* Nanoscale broadband transmission lines for spin qubit control. *Nanotechnology* **24**, 015202 (2013).
26. Pines, D., Bardeen, J. & Slichter, C. P. Nuclear polarization and impurity-state spin relaxation processes in silicon. *Phys. Rev.* **106**, 489–498 (1957).
27. Rehner, L., Hoehne, F., Stutzmann, M. & Brandt, M. S. Nuclear spins of ionized phosphorus donors in silicon. *Phys. Rev. Lett.* **108**, 027602 (2012).
28. Witzel, W. M., Carroll, M. S., Cywinski, L. & Das Sarma, S. Quantum decoherence of the central spin in a sparse system of dipolar coupled spins. *Phys. Rev. B* **86**, 035452 (2012).
29. Mehring, M., Mende, J. & Scherer, W. Entanglement between an electron and a nuclear spin 1/2. *Phys. Rev. Lett.* **90**, 153001 (2003).
30. Simmons, S. *et al.* Entanglement in a solid-state spin ensemble. *Nature* **470**, 69–72 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. P. Starrett, D. Barber, C. Y. Yang and R. Szymanski for technical assistance and A. Laucht, R. Kalla and J. Muhonen for discussions. This research was funded by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project number CE11E0096) and the US Army Research Office (W911NF-13-1-0024). We acknowledge support from the Australian National Fabrication Facility.

Author Contributions K.Y.T. and W.H.L. fabricated the device; D.N.J. designed the P implantation experiments; J.J.P., K.Y.T., J.J.L.M., J.P.D. and F.A.Z. performed the measurements; J.J.P., A.M., A.S.D. and J.J.L.M. designed the experiments and discussed the results; J.J.P. and A.M. analysed the data; and J.J.P. and A.M. wrote the manuscript with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (a.morello@unsw.edu.au).

Unexpected strain-stiffening in crystalline solids

Chao Jiang^{1†} & Srivilliputhur G. Srinivasan²

Strain-stiffening—an increase in material stiffness at large strains—is a vital mechanism by which many soft biological materials thwart excessive deformation to protect tissue integrity^{1–3}. Understanding the fundamental science of strain-stiffening and incorporating this concept into the design of metals and ceramics for advanced applications is an attractive prospect. Using cementite (Fe₃C) and aluminium borocarbide (Al₃BC₃) as prototypes, here we show via quantum-mechanical calculations that strain-stiffening also occurs, surprisingly, in simple inorganic crystalline solids and confers exceptionally high strengths to these two solids, which have anomalously low resistance to deformation near equilibrium. For Fe₃C and Al₃BC₃, their ideal shear strength to shear modulus ratios attain remarkably high values of 1.14 and 1.34 along the (010)[001] and (0001)[0110] slip systems, respectively. These values are more than seven times larger than the original Frenkel value of $1/2\pi$ (refs 4, 5) and are the highest yet reported for crystalline solids. The extraordinary stiffening of Fe₃C arises from the strain-induced reversible ‘cross-linking’ between weakly coupled edge- and corner-sharing Fe₆C slabs. This new bond formation creates a strong, three-dimensional covalent bond network that resists large shear deformation. Unlike Fe₃C, no new bond forms in Al₃BC₃ but stiffening still occurs because strong repulsion between Al and B in a compressed Al–B bond unsettles the existing covalent bond network. These discoveries challenge the conventional wisdom that large shear modulus is a reliable predictor of hardness and strength of materials^{4–7}, and provide new lessons for materials selection and design.

Vigorous research into the atomistic origin of strength of materials continues unabated. It is well known that most materials are weakened and ultimately break the more they are strained. At low strains, stress in a material is linearly proportional to its strain and the constant of proportionality is the modulus of elasticity or a material’s resistance to elastic deformation. Many structurally complex biological materials^{1–3} are notable exceptions to this rule. Although such materials are initially soft they rapidly stiffen at larger strains. This nonlinear elastic response is vital for the physiological function of tissues like skin and blood vessels, although the underlying mechanism is still not fully understood.

Here we use quantum-mechanical calculations based on density functional theory (DFT) to show and elucidate a surprising strain-induced stiffening phenomenon in two structurally simple inorganic crystalline solids: orthorhombic Fe₃C and hexagonal Al₃BC₃. Owing to strain-stiffening, both materials possess high ideal shear strengths despite their extremely soft elastic shear moduli c_{44} . The classical Frenkel model^{4,5} estimates the maximum attainable shear strength τ_{\max} of a perfect crystal, the so-called ideal shear strength, to be $\tau_{\max} \approx G/2\pi$ —about a tenth of its shear modulus G calculated at small strains. However, surprisingly, the τ_{\max}/G ratios for both Fe₃C and Al₃BC₃ are greater than 1.1.

Steels set the ‘gold standard’ for strong materials and continue to be the most important and well-studied structural material in the three millennia since the start of the Iron Age. We now know that carbon steels are primarily strengthened by the precipitation of excess carbon as hard Fe₃C carbide (Fig. 1a) instead of soft graphite. The fraction and morphology of cementite, for example in pearlite and bainite, sensitively

control the strength and ductility of carbon steels. Cementite has also been proposed to be a major constituent of Earth’s core⁸.

In contrast to its immense technological and scientific importance, the properties of cementite are still poorly understood owing to difficulties in growing crystals large enough and pure enough for experimentation. An interesting finding of our previous DFT study⁹ is the extreme elastic anisotropy of Fe₃C, due to its anomalously low shear modulus c_{44} , which is about tenfold smaller than c_{55} and c_{66} , and over twofold less than the c_{44} of soft aluminium¹⁰. Experimental validation of our DFT predictions comes from the recent X-ray diffraction stress measurements¹¹. This extremely low c_{44} may be indicative of mechanical instability and weak resistance to shear in the (010)[001] slip system, which contradicts the observed predominance and stability of cementite in steels¹², as well as its high hardness (Vickers hardness, ~ 10 GPa; ref. 13). The recent nanoindentation experiments¹⁴ that show a significantly smaller anisotropy in the indentation modulus of cementite compared to our DFT predictions⁹ add further intrigue to this puzzle. We note that extremely low c_{44} and large elastic anisotropy may also be found in simple cubic polonium and in ionic compounds with NaCl structure¹⁵. A detailed discussion of stability of crystalline solids may be found in a recent review¹⁶.

To resolve these experimental–theoretical controversies, here we use DFT calculations to predict the stress–strain relations of ideal, defect-free Fe₃C crystal to the limit of its structural stability^{17–21}. These calculations quantify the ideal strength of cementite, which sets an upper bound on the strength of the material and can be compared with results from nanoindentation experiments. Surprisingly, they reveal an intriguing reversible, nonlinear elastic response to applied shear strain that we attribute to dramatic changes in bonding characteristics at large strains. These new insights are important inputs to physics-informed deformation models. Details of our computational methods and discussions of the possible effects of phonon and magnetic instabilities are given in the Supplementary Information.

Figure 2a shows the stress–strain relations for Fe₃C under uniaxial tension along various crystallographic directions. For [100] and [001] tension, the stress increases monotonically with strain and then drops rapidly above critical strains of 16% and 23%, respectively. In contrast, for the [010] tension, there exists a noticeable stress plateau above 23% strain. The ideal tensile strengths, defined as the first maximum in a stress–strain curve, are 22.4 GPa, 19.5 GPa and 31.3 GPa along the [100], [010] and [001] directions respectively. The initial slopes of the stress–strain curves give Young’s moduli $E_{[100]}$, $E_{[010]}$, and $E_{[001]}$ of 283.4 GPa, 245.4 GPa and 221.4 GPa, respectively. These values agree well with corresponding moduli of 284.0 GPa, 245.8 GPa and 220.0 GPa calculated from single-crystal elastic constants⁹.

To understand the origin of the stress plateau under [010] tension, but absent under [100] or [001] tension, we examine its orthorhombic crystal structure (space group $Pnma$), as shown in Fig. 1a. The 12 Fe atoms occupy special (Wyckoff position 4c) and general (Wyckoff position 8d) sites. The carbon atoms occupy only special 4c sites. The fundamental building blocks of cementite are the Fe₆C trigonal prisms, which form a two-dimensional slab parallel to the (010) plane via edge- and corner-sharing (Fig. 1a). Interestingly, while there exists

¹Structure/Property Relations Group (MST-8), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²Department of Materials Science and Engineering, University of North Texas, Denton, Texas 76203, USA. [†]Present address: Department of Materials Science and Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA.

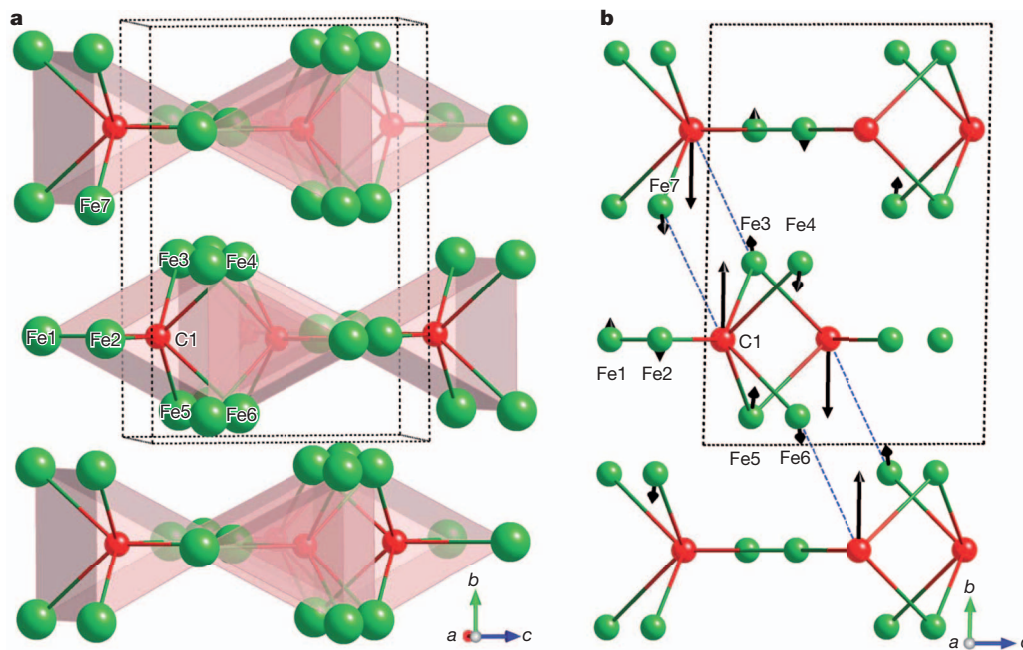


Figure 1 | Crystal structure of cementite. **a**, At equilibrium. Green and red spheres represent Fe and C atoms, respectively. The six Fe atoms that are the first nearest neighbours of a C atom are labelled Fe1 to Fe6. **b**, Inhomogeneous internal atomic relaxations in response to a 2% (010)[001] shear strain. The arrows give the directions and relative magnitudes of atomic displacements.

The blue dashed lines indicate Fe7–C1 bonds that will form at large shear strains, but are absent in the equilibrium structure. The black dashed lines represent the primitive unit cell of cementite; pink shading is used to highlight the Fe_6C trigonal prisms as the fundamental building blocks of cementite.

a mixture of covalent Fe–C (due to hybridization between C $2p$ and Fe $3d$ states) and metallic Fe–Fe bonding within each slab, Fe–Fe metallic bonds couple adjacent slabs⁹. Uniaxial tension along the [010] direction predominantly elongates the interslab Fe–Fe bonds without appreciably stretching the stronger Fe–C bonds (Fig. 2b). This produces a ‘metal-like’ deformation of cementite under [010] tensile loading. In contrast, the abrupt snapping of strong covalent Fe–C bonds within Fe_6C trigonal prisms destabilizes cementite under [100] and [001] tension and creates a steep stress drop at the critical strain.

Figure 2c shows the stress–strain curves for Fe_3C under shear deformation on various slip systems. The initial slopes of these curves give c_{44} , c_{55} and c_{66} shear moduli of 12.2 GPa, 133.8 GPa and 136.9 GPa, respectively, in excellent agreement with our previous study⁹ ($c_{44} = 15$ GPa, $c_{55} = 134$ GPa and $c_{66} = 134$ GPa). A particularly surprising finding is the strain-induced stiffening of cementite. Because

the c_{44} elastic shear modulus associated with the (010)[001] slip system of cementite is exceptionally low, one would expect its shear strength to be correspondingly weak. However, although the slope of the stress–strain curve under (010)[001] shear is rather low for small strains, it rapidly stiffens with increasing shear strain. While strain-stiffening has been observed in many biological materials^{1–3} and elastomers²², this is the first time such a phenomenon has been predicted in a simple crystalline solid, to the best of our knowledge.

Remarkably, despite the fact that c_{44} is only 1/10 of c_{66} , the ideal shear strength of 13.9 GPa for the (010)[001] slip system exceeds the 11.9 GPa obtained for the (010)[100] system, and is only slightly smaller than the 17.1 GPa for the (100)[001] system. Thus, owing to strain stiffening, cementite exhibits an unusual combination of extremely high elastic shear anisotropy and almost isotropic shear strength. This is in striking contrast with, for example, tungsten, which is elastically isotropic at low

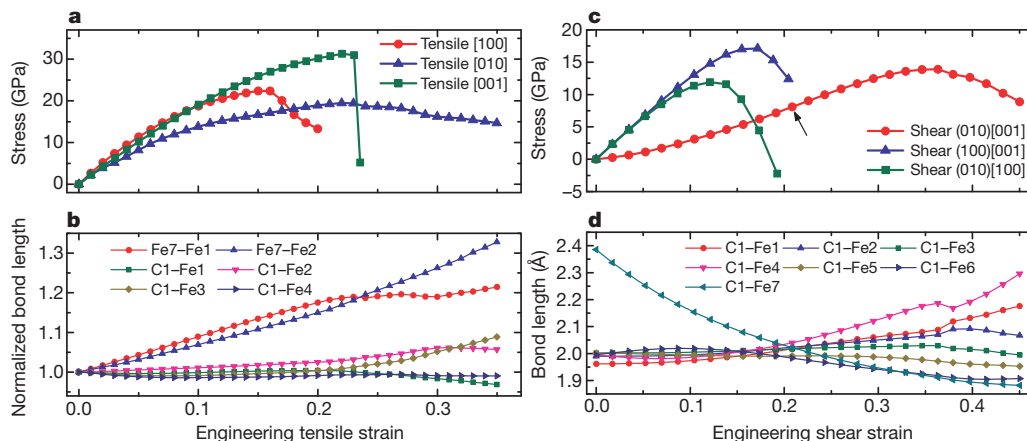


Figure 2 | Fe_3C under tensile and shear deformations. **a**, Stress–strain curves under [100], [010] and [001] uniaxial tension. **b**, Variations of Fe–Fe and Fe–C bond lengths with strain under uniaxial [010] loading. By symmetry, the C1–Fe5 and C1–Fe6 bond lengths are equal to those of C1–Fe3 and of C1–Fe4,

respectively. Bond lengths are normalized by their equilibrium values. **c**, Stress–strain relations along various shear paths. For the (001)[001] shear, an arrow marks the inflection point of the stress–strain curve. **d**, Changes of Fe–C bond lengths as a function of applied (010)[001] shear strain.

strains and exhibits a high anisotropy in tensile strengths²³. This insight resolves the puzzling inconsistencies between some recent experiments: although X-ray diffraction stress measurements¹¹ at small strains indicated extreme elastic anisotropy of Fe₃C in agreement with our earlier DFT predictions⁹, indentation moduli obtained from nanoindentation measurements at large strains are much more isotropic¹⁴. For the (010)[001] slip system in cementite, the ratio of the ideal shear strength to the associated elastic shear modulus (1.14) is anomalously large—about sevenfold larger than the Frenkel value. Note that τ_{\max}/G ratios are normally between 0.04 and 0.24 (ref. 5).

In nanoindentation experiments, owing to the small (and hence potentially dislocation-free) volumes probed, it is generally believed that the shear stress at the onset of plasticity can approach the theoretical shear strength of an ideal, defect-free material²⁴. Furthermore, *in situ* nanoindentation experiments²⁵ in a transmission electron microscope provided direct evidence that a submicrometre grain of aluminium plastically deformed by nanoindentation to a dislocation density of $\sim 10^{14} \text{ m}^{-2}$ can also support shear stresses close to the theoretical shear strength. This result is contrary to earlier assumptions that a dislocation-free volume is necessary to achieve shear stresses near the theoretical shear strength of the material. For Fe₃C, the measured

critical shear stress for incipient plasticity is nearly independent of crystal orientation and its mean value (11.9 GPa; ref. 14) is indeed close to our calculated ideal shear strengths (11.9–17.1 GPa). Remarkably, in nanoindentation experiments¹⁴, the critical resolved shear stress to shear modulus ratio for the (010)[001] slip system was also found to be abnormally large (~ 1.0), which is again in good agreement with our present findings.

The cementite crystal contains many internal atomic degrees of freedom not fixed by space group symmetry. Upon straining, these atomic degrees of freedom readjust to minimize the strain energy. Furthermore, new atomic degrees of freedom can also emerge due to strain-induced reduction in crystal symmetry⁹. For example, (010)[001] shear deformation reduces the symmetry of cementite from orthorhombic to monoclinic (space group $P2_1/c$). This allows the y -coordinates of Fe and C atoms at special positions, which are initially fixed by symmetry, to relax. As shown in Fig. 1b, C atoms respond to the (010)[001] shear by moving mainly along the [010] direction towards a neighbouring slab. As a consequence, the distance between C1 and Fe7 atoms in two adjacent slabs decreases progressively with increasing shear strain and becomes comparable to the average Fe–C bond length in equilibrium Fe₃C (1.99 Å) at about 21% strain (Fig. 2d). Interestingly, this strain also

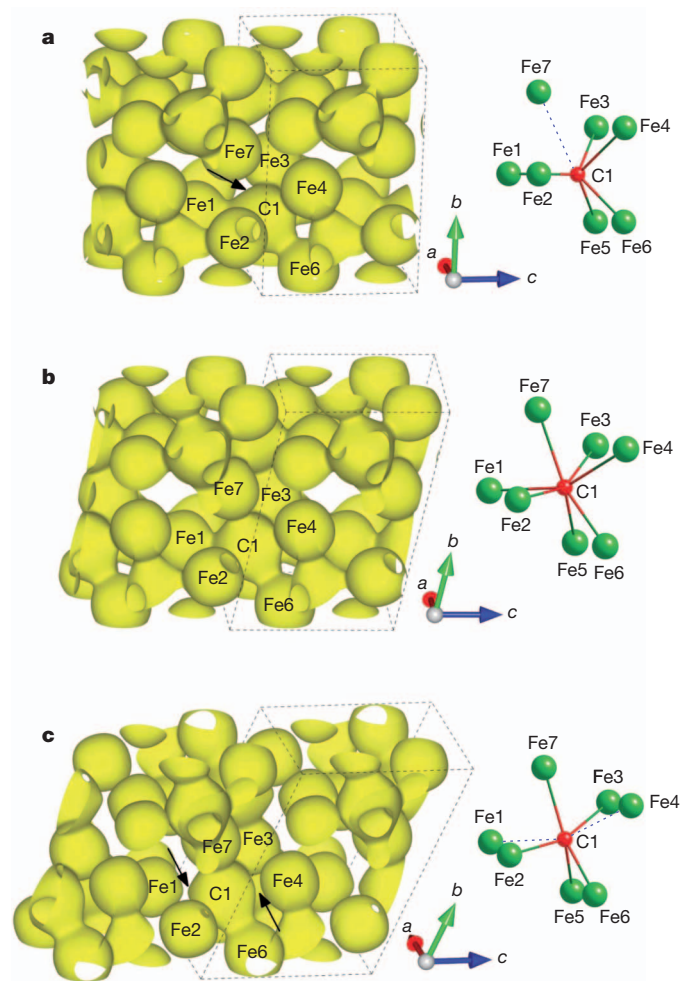


Figure 3 | Valence charge density in cementite. Isosurfaces at $\rho = 0.5e \text{ Å}^{-3}$ along the (010)[001] shear path are shown for strains $\gamma = 0$ (a), $\gamma = 0.21$ (at the inflection point of the stress-strain curve; b), and $\gamma = 0.38$ (just after lattice instability; c). Ball-and-stick images in the right column illustrate the variation of the local coordination of C atoms with shear strain. The arrows point to regions of charge depletion between Fe and C atoms, indicating broken bonds (blue dashed lines in the ball and stick images). The grey dashed lines represent the primitive unit cell.

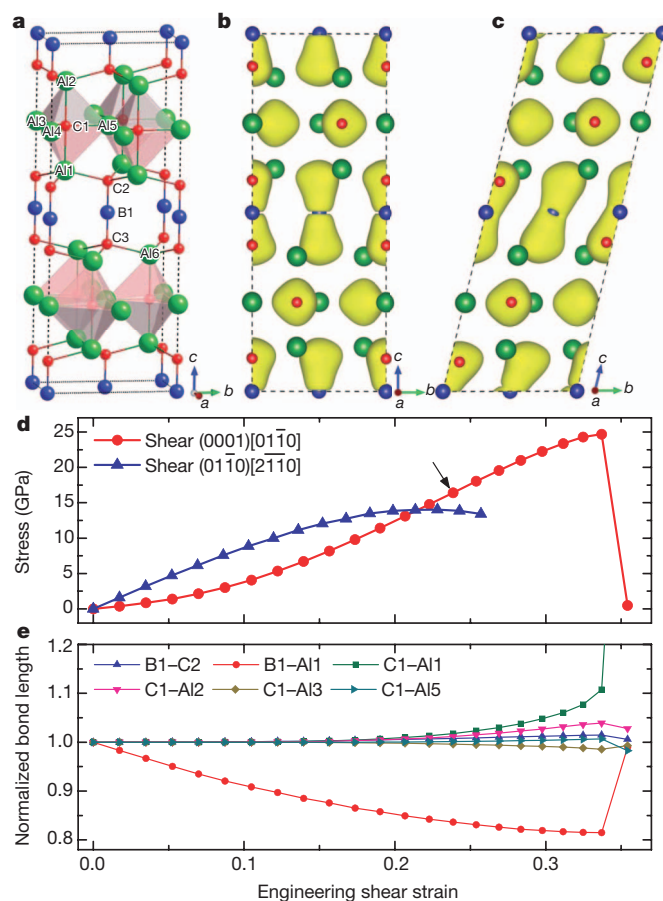


Figure 4 | Shear deformation of Al₃BC₃ along two different slip systems. a, The crystal structure of Al₃BC₃. The a , b and c axes are parallel to the $[2\bar{1}10]$, $[01\bar{1}0]$ and $[0001]$ directions of the primitive hexagonal unit cell, respectively. Green, blue and red spheres represent Al, B and C atoms, respectively. Isosurfaces of valence charge density at $\rho = 0.5e \text{ Å}^{-3}$ along the (0001)[0110] shear path are shown for strains $\gamma = 0$ (b) and $\gamma = 0.24$ (at the inflection point of the stress-strain curve; c). d, Stress-strain relation under (0001)[0110] and (0110)[2110] shear. For (0001)[0110] shear, the arrow marks the inflection point of the stress-strain curve. e, Variations of normalized bond lengths with (0001)[0110] shear strain. By symmetry, the B1–C3, B1–Al6 and C1–Al4 bond lengths are equal to those of B1–C2, B1–Al1 and C1–Al3, respectively.

coincides with the inflection point of the stress–strain curve separating the strain-stiffening and strain-softening regimes.

At this 21% strain, each C atom now forms seven strong covalent bonds with neighbouring Fe atoms, as can be seen from Fig. 3b. A three-dimensional covalent bonding network is created in cementite because the new Fe–C covalent bonds ‘cross-link’ adjacent slabs. This three-dimensional network of Fe–C bonds, which is absent in the equilibrium cementite structure (Fig. 3a), effectively resists shear and causes the unexpected strain stiffening in cementite. When sheared above 21% strain, the slope of the stress–strain curve begins to decrease and the cementite lattice softens and becomes unstable at a critical strain of 36%, when both C1–Fe1 and C1–Fe4 bonds break. This is evident from the charge depletion in the bond region between corresponding atoms (Fig. 3c). Our Mulliken overlap population analysis, which gives a direct measurement of bond covalency, provides a more quantitative evidence of new bond formation and bond breaking in Fe₃C under (010)[001] shear (Supplementary Fig. 3).

Our results naturally lead to the question of whether strain-stiffening is present in other inorganic materials. To this end, we have studied the shear deformation of hexagonal Al₃BC₃ (space group *P*_{6₃/mmc), another layered material (Fig. 4a) with high hardness (Vickers hardness, ~14 GPa; ref. 26) but anomalously low shear modulus²⁷. Despite its low elastic shear modulus *c*₄₄ (16 GPa from ref. 27 and 18.4 GPa from our study), the Al₃BC₃ lattice also stiffens steeply under (0001)[01 $\bar{1}$ 0] shear strain, giving surprisingly high ideal shear strength of 24.7 GPa (Fig. 4d). The resulting τ_{max}/G ratio (1.34) is the highest ever known for a crystalline solid. In contrast, the ideal shear strength along the (01 $\bar{1}$ 0)[2 $\bar{1}$ $\bar{1}$ 0] slip system is only 14.0 GPa, although the related elastic shear modulus *c*₆₆ (95 GPa from ref. 27 and 92.6 GPa from our study) is over fivefold larger than *c*₄₄.}

Unlike cementite, no new bonds form during shearing of Al₃BC₃ (Fig. 4b and c). Instead, the atomistic origin of strain stiffening in Al₃BC₃ is the strong repulsion between Al and B atoms due to severe squeezing of the B1–Al1 and the B1–Al6 bonds by the tilting of the rigid linear C–B–C units (Fig. 4e). Eventually, Al₃BC₃ fails under (0001)[01 $\bar{1}$ 0] shear owing to the complete rupture of the C1–Al1 bond within the Al₅C trigonal bipyramid, after which the equilibrium B1–Al1 and B1–Al6 bond lengths are restored. Note that whereas cementite is ferromagnetic and a metallic conductor⁹, Al₃BC₃ is a non-magnetic semiconductor²⁷. We anticipate that strain-stiffening may operate in other low-symmetry, layered materials with atomic degrees of freedom for internal relaxations.

In conclusion, we have discovered that strain-stiffening, a popular phenomenon in many biological materials and elastomers^{1–3,22}, also unexpectedly occurs in simple crystalline solids. We have identified two different strain-stiffening mechanisms operating in inorganic solids that confer surprisingly high strength to materials with extremely low elastic constants. Our work highlights the untapped possibilities and richness of the strain-stiffening phenomenon in solids, and challenges the current philosophy of designing (super)hard materials largely on the basis of strong elastic properties^{4–7,28–30}.

METHODS SUMMARY

Ab initio ideal strength calculations are performed using the all-electron projector augmented wave method within the generalized gradient approximation of Perdew, Burke and Ernzerhof (ref. 2 in the Supplementary Information), as implemented in the Vienna *ab initio* simulation package (ref. 1 in the Supplementary Information). Spin-polarized calculations are performed to account for the ferromagnetic nature of Fe₃C. A high plane-wave cutoff energy of 500 eV and dense *k*-point sampling are employed to guarantee high numerical accuracy. To provide a more quantitative bonding picture of Fe₃C, we have performed Mulliken bond overlap population analysis using the CASTEP code (ref. 5 in the Supplementary Information). Further details are given in the Supplementary Information.

Received 26 August 2012; accepted 6 February 2013.

Published online 10 April 2013.

- Storm, C., Pastore, J. J., Mackintosh, F. C., Lubensky, T. C. & Janmey, P. A. Nonlinear elasticity in biological gels. *Nature* **435**, 191–194 (2005).
- Shadwick, R. E. Mechanical design in arteries. *J. Exp. Biol.* **202**, 3305–3313 (1999).
- Meyers, M. & Chawla, K. *Mechanical Behavior of Materials* Ch. 2 (Cambridge Univ. Press, 2008).
- Frenkel, J. Zur theorie der elastizitätsgrenze und der festigkeit kristallinischer körper. *Z. Phys.* **37**, 572–609 (1926).
- Kelly, A. *Strong Solids* Ch. 1 (Clarendon Press, 1973).
- Teter, D. M. Computational alchemy: the search for new superhard materials. *Mater. Res. Bull.* **23**, 22–27 (1998).
- Jhi, S. H., Ihm, J., Louie, S. G. & Cohen, M. L. Electronic mechanism of hardness enhancement in transition-metal carbonitrides. *Nature* **399**, 132–134 (1999).
- Wood, B. J. Carbon in the core. *Earth Planet. Sci. Lett.* **117**, 593–607 (1993).
- Jiang, C., Srinivasan, S. G., Caro, A. & Maloy, S. A. Structural, elastic, and electronic properties of Fe₃C from first principles. *J. Appl. Phys.* **103**, 043502 (2008).
- Wang, H. & Li, M. *Ab initio* calculations of second-, third-, and fourth-order elastic constants for single crystals. *Phys. Rev. B* **79**, 224102 (2009).
- Nikolussi, N. et al. Extreme elastic anisotropy of cementite, Fe₃C: first-principles calculations and experimental evidence. *Scr. Mater.* **59**, 814–817 (2008).
- Fang, C. M., Sluiter, M. H. F., van Huis, M. A., Ande, C. K. & Zandbergen, H. W. Origin of predominance of cementite among iron carbides in steel at elevated temperature. *Phys. Rev. Lett.* **105**, 055503 (2010).
- Umemoto, M., Liu, Z. G., Masuyama, K. & Tsuchiy, K. Influence of alloy additions on production and properties of bulk cementite. *Scr. Mater.* **45**, 391–397 (2001).
- Alkorta, J. & Sevillano, J. G. Assessment of elastic anisotropy and incipient plasticity in Fe₃C by nanoindentation. *J. Mater. Res.* **27**, 45–52 (2012).
- Legut, D., Friak, M. & Sob, M. Phase stability, elasticity, and theoretical strength of polonium from first principles. *Phys. Rev. B* **81**, 214118 (2010).
- Grimvall, G., Magyari-Kope, B., Ozolins, V. & Persson, K. A. Lattice instabilities in metallic elements. *Rev. Mod. Phys.* **84**, 945–986 (2012).
- Roundy, D., Krenn, C. R., Cohen, M. L. & Morris, J. W. Ideal shear strengths of fcc aluminum and copper. *Phys. Rev. Lett.* **82**, 2713–2716 (1999).
- Telling, R. H., Pickard, C. J., Payne, M. C. & Field, J. E. Theoretical strength and cleavage of diamond. *Phys. Rev. Lett.* **84**, 5160–5163 (2000).
- Chacham, H. & Kleinman, L. Instabilities in diamond under high shear stress. *Phys. Rev. Lett.* **85**, 4904–4907 (2000).
- Jhi, S. H., Louie, S. G., Cohen, M. L. & Morris, J. W. Mechanical instability and ideal shear strength of transition metal carbides and nitrides. *Phys. Rev. Lett.* **87**, 075503 (2001).
- Ogata, S., Li, J. & Yip, S. Ideal pure shear strength of aluminum and copper. *Science* **298**, 807–811 (2002).
- Zhao, H. E. & Suo, Z. G. Theory of dielectric elastomers capable of giant deformation of actuation. *Phys. Rev. Lett.* **104**, 178302 (2010).
- Sob, M., Wang, L. G. & Vitek, V. Theoretical tensile stress in tungsten single crystals by full-potential first-principles calculations. *Mater. Sci. Eng. A* **234–236**, 1075–1078 (1997).
- Krenn, C. R., Roundy, D., Cohen, M. L., Chrzan, D. C. & Morris, J. W. Connecting atomistic and experimental estimates of ideal strength. *Phys. Rev. B* **65**, 134111 (2002).
- Minor, A. M. et al. A new view of the onset of plasticity during the nanoindentation of aluminium. *Nature Mater.* **5**, 697–702 (2006).
- Lee, S. H. et al. Chemical composition and microstructure of Al₃BC₃ prepared by different densification methods. *J. Eur. Ceram. Soc.* **30**, 1015–1020 (2010).
- Wang, J. Y., Zhou, Y. C., Liao, T. & Lin, Z. J. First-principles prediction of low shear-strain resistance of Al₃BC₃: a metal borocarbide containing short linear BC₂ units. *Appl. Phys. Lett.* **89**, 021917 (2006).
- Teter, D. M. & Hemley, R. J. Low-compressibility carbon nitrides. *Science* **271**, 53–55 (1996).
- Dubrovinsky, L. S. et al. The hardest known oxide. *Nature* **410**, 653–654 (2001).
- Kaner, R. B., Gilman, J. & Tolbert, S. H. Designing superhard materials. *Science* **308**, 1268–1269 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements C.J. acknowledges the support of a Director's Fellowship at Los Alamos National Laboratory (LANL), where a systematic study of cementite was conceived and initiated. S.G.S. acknowledges support from the National Science Foundation (grant number 0846444). We also thank J. Wills, M. I. Baskes, A. Caro, A. Misra, S. Maloy, A. Srivastava, V. Vitek and S. Ranganathan for their discussions.

Author Contributions C.J. and S.G.S. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J. (chaopu@gmail.com) and S.G.S. (srinivasan.srivilliputhur@unt.edu).

Mapping molecular motions leading to charge delocalization with ultrabright electrons

Meng Gao^{1,2*}, Cheng Lu¹, Hubert Jean-Ruel^{1,2}, Lai Chung Liu^{1,2}, Alexander Marx², Ken Onda^{3,4}, Shin-ya Koshihara^{5,6}, Yoshiaki Nakano⁷, Xiangfeng Shao^{7†}, Takaaki Hiramatsu⁸, Gunzi Saito⁸, Hideki Yamochi⁷, Ryan R. Cooney^{1,2}, Gustavo Moriena^{1,2}, Germán Sciaini^{1,2*} & R. J. Dwayne Miller^{1,2}

Ultrafast processes can now be studied with the combined atomic spatial resolution of diffraction methods and the temporal resolution of femtosecond optical spectroscopy by using femtosecond pulses of electrons^{1–14} or hard X-rays^{15–19} as structural probes. However, it is challenging to apply these methods to organic materials, which have weak scattering centres, thermal lability, and poor heat conduction. These characteristics mean that the source needs to be extremely bright to enable us to obtain high-quality diffraction data before cumulative heating effects from the laser excitation either degrade the sample or mask the structural dynamics²⁰. Here we show that a recently developed, ultrabright femtosecond electron source^{7–9} makes it possible to monitor the molecular motions in the organic salt (EDO-TTF)₂PF₆ as it undergoes its photo-induced insulator-to-metal phase transition^{21–24}. After the ultrafast laser excitation, we record time-delayed diffraction patterns that allow us to identify hundreds of Bragg reflections with which to map the structural evolution of the system. The data and supporting model calculations indicate the formation of a transient intermediate structure in the early stage of charge delocalization (less than five picoseconds), and reveal that the molecular motions driving its formation are distinct from those that, assisted by thermal relaxation, convert the system into a metallic state on the hundred-picosecond timescale. These findings establish the potential of ultrabright femtosecond electron sources^{7–14} for probing the primary processes governing structural dynamics with atomic resolution in labile systems relevant to chemistry and biology.

(EDO-TTF)₂PF₆ (where EDO-TTF is ethylenedioxytetrathiafulvalene) is a quasi-one-dimensional, 3/4-band-filled charge-transfer organic salt that undergoes a thermally induced insulator-to-metal phase transition at a critical temperature of $T_c \approx 280$ K (ref. 21) and also a highly efficient and ultrafast photo-induced phase transition^{22–24}. Its electronic structure resembles that of Bechgaard salts, which provided the first organic superconductor, (TMTSF)₂PF₆ (see ref. 25). The origin of its insulator-to-metal phase transition involves a variety of collective phenomena²³: Peierls-and-Holstein-type electron-phonon²⁶ and anti-ferromagnetic interactions²⁷, that coexist with order-disorder^{21,24} charge localization, and long-range Coulombic interactions have an important role in charge disproportionation²⁸. In its metallic or high-temperature (HT) phase, electron donor EDO-TTF molecules (D) form columns of cations that are separated by sheets of acceptor PF₆ anions. The distribution of positive charges among EDO-TTF molecules along the stacking direction is represented by $(D^{+0.5}D^{+0.5}D^{+0.5}D^{+0.5})$ as shown in Fig. 1 (top right)^{21–24}. The high positive charge mobility along the cation stacks confers metallic properties at room temperature. The holes localize below T_c and endow the low-temperature (LT) phase with insulating

properties. In the LT phase, the EDO-TTF molecules present a charge-ordered^{21–24} state $(D^{+1}D^{+0}D^{+0}D^{+1})$, in which a large bending of the neutral moieties promotes the doubling of the unit cell akin to a Peierls mechanism in a half-band-filled system, as shown in Fig. 1 (top left). Model calculations predict that vertical photo-excitation via the second charge-transfer band²⁶ leads to a localized $\{D^{+2}D^{+0}D^{+0}D^{+0}\}$ excited state. Time-resolved optical reflectivity measurements in combination with theoretical modelling indicate²³ that this initial excited state evolves in less than 100 fs into a $(D^{+1}D^{+0}D^{+1}D^{+0})$ charge-disproportionate state, which has a lifetime of about 4 ps. Time-resolved optical studies in the near- and mid-infrared region identify large charge fluctuations as the driving process that finally leads to the complete randomization and melting of the charge order after about 100 ps (ref. 24).

To probe the structural evolution in (EDO-TTF)₂PF₆ directly during its photo-induced insulator-to-metal phase transition, we performed femtosecond electron diffraction (FED) studies employing a recently developed ultrabright femtosecond electron source^{7–9} that can

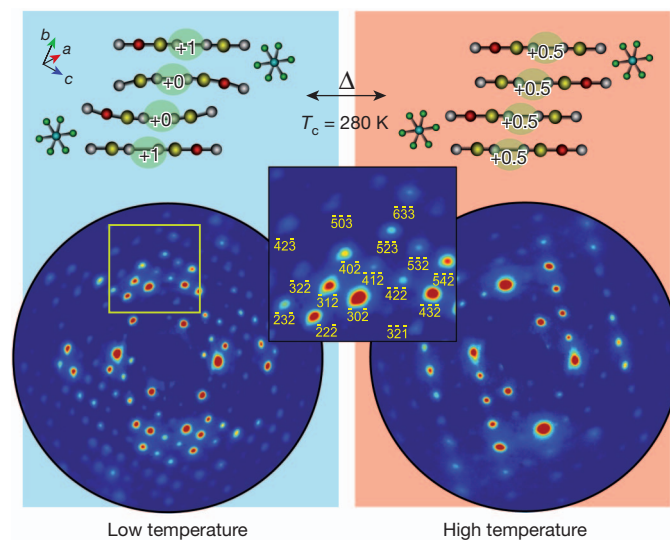


Figure 1 | Insulator-to-metal first-order phase transition in (EDO-TTF)₂PF₆. Top panels, illustration of the molecular and electronic changes associated with the thermal insulator-to-metal phase transition. Bottom panels, diffraction patterns for the LT and HT phases, obtained at 230 K and 295 K, respectively. The inset shows the assigned Miller indices (h, k, l). The symmetry breaking (cell doubling) corresponds to peaks indexed with $k = 2n + 1$ in the LT phase. For details about the definition of unit cell axes, see Supplementary Information section 6.1.

¹Departments of Chemistry and Physics, University of Toronto, Toronto, Ontario M5S 3H6, Canada. ²Max Planck Research Department for Structural Dynamics, Department of Physics, University of Hamburg, Hamburg Center for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany. ³Interactive Research Center of Science, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8502, Japan. ⁴PRESTO, Japan Science and Technology Agency, Honcho, Kawaguchi 332-0012, Japan. ⁵Department of Chemistry and Materials Science, Tokyo Institute of Technology, Ōokayama, Meguro-ku, Tokyo 152-8551, Japan. ⁶CREST, Japan Science and Technology Agency (JST), 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-8666, Japan. ⁷Research Center for Low Temperature and Materials Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan. ⁸Faculty of Agriculture, Meijo University, Shiogamaguchi 1-501 Tempaku-ku, Nagoya 468-8502, Japan. [†]Present address: State Key Laboratory of Applied Organic Chemistry, Lanzhou University, 222 Tianshui South Road, Lanzhou 730000, Gansu, China.

*These authors contributed equally to this work.

deliver high-quality diffraction from low-atomic-number materials even at the very low repetition rate of only 10 Hz used to avoid the adverse effects caused by cumulative heating. FED experiments were performed in transmission mode, with the electron beam propagating nearly perpendicular to the stacking direction of the EDO-TTF moieties to ensure high sensitivity to the most relevant molecular motions, that is, those associated with the dynamics of the EDO-TTF moieties and PF_6 counter-ions. Samples were kept at 230 K (for the LT phase) and photo-excited by 60-fs, 800-nm light while being monitored by ultra-bright femtosecond electron pulses, with 200 or more reflections enabling us to track time-dependent changes beyond the noise level. See the Supplementary Information for detailed information about methodology (sections 1–3), sample preparation (section 5), sample orientation (sections 6.1 and 6.4), and control experiments (section 4).

Figure 1 depicts the molecular arrangement of EDO-TTF moieties with their respective charge distributions for the LT and HT phases, along with the corresponding electron diffraction patterns. The appearance of a new family of peaks in the diffraction pattern of the LT phase is indicative of a cell-doubling effect, with the clear difference between the diffraction patterns of the LT and HT phases confirming that the crystal orientation allows the observation of the relevant structural changes.

We monitor the structural evolution of the system by collecting diffraction patterns at different time delays after photoexcitation and using the pattern of the initial LT phase as a reference, with the resulting difference patterns shown in Fig. 2a. For comparison, Fig. 2b shows the difference between the diffraction patterns of the HT and LT phases (shown in Fig. 1). Figure 2c gives the temporal evolution of normalized intensity changes for some selected Bragg reflections, which allow separation of the observed photo-induced structural dynamics into a ‘fast’ component in the subpicosecond-to-picosecond

range and a ‘slow’ structural change proceeding on the 100-ps time-scale. The fast dynamics can be explained as femtosecond optical excitation altering the charge distribution, modifying the inter- and intramolecular forces and driving coherent molecular motions. The slow dynamics are attributed to uncorrelated motions that provide an ensemble-averaged picture of the structural evolution, that is, a thermal relaxation process that brings the system to a metallic HT-type state²⁴. The transfer of heat to the lattice manifests itself as slow oscillations of diffracted intensities (see right panel in Fig. 2c) which are probably introduced through the generation of strain waves²⁹. The most noteworthy feature is the plateau behaviour observed for time delays from about 3 ps to 10 ps (left panel of Fig. 2c), which reflects a transient intermediate structure (TIS). Given that optical studies indicate large charge fluctuations²³ prevailing in this time span²⁴, it seems plausible that the effective intermolecular forces are screened and the structure therefore transiently locked.

To obtain a time-dependent map of the relevant molecular motions driving the insulator-to-metal phase transition and the formation of the TIS, we developed a structural refinement algorithm based on a parameterized molecular model (see equations (4), (6), (7) and (8) in Supplementary Information section 7). The model makes use of the LT and HT structures as previously determined by X-ray diffraction²¹ and considers about 40 Bragg reflections in the refinement calculations. The selection of reflections was based on their proximity to the Ewald sphere (Bragg condition) and the necessity of removing overlapped peaks.

The entire data set of diffraction images was reduced to the square root of intensities, that is, structure factor amplitudes $|F_{\text{exp}}(k, t)|$ (besides a k -dependent constant), where k denotes the scattering vector (or Miller indices) and t is a given time delay. Given that optical excitation causes only a small fraction of unit cells (η_{exc}) to undergo

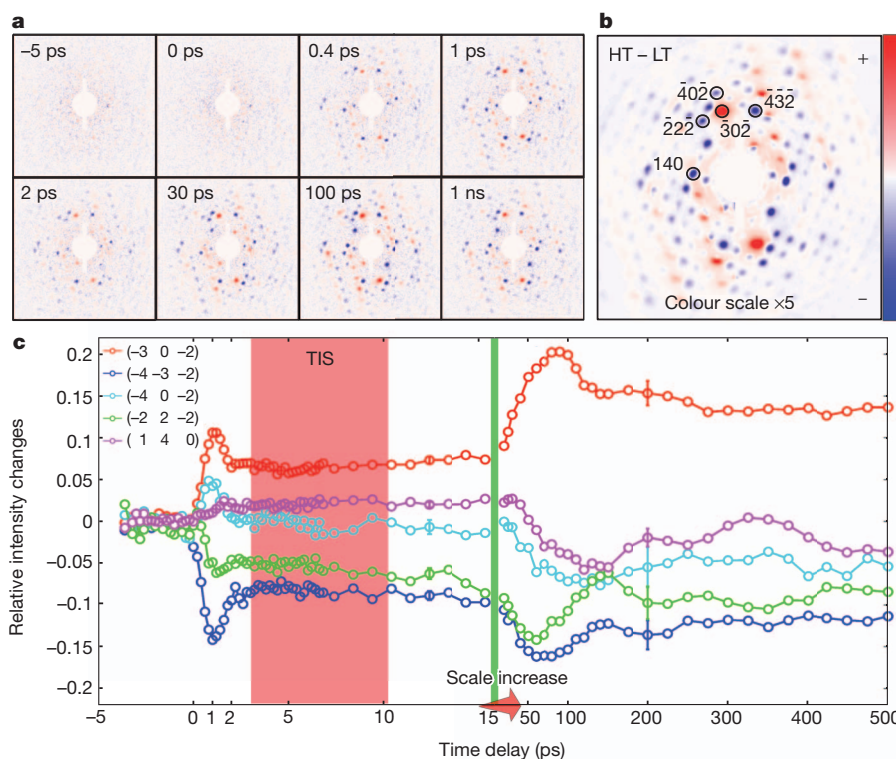


Figure 2 | Photo-induced time-dependent structural changes monitored by ultrabright FED. **a**, Differences between the diffraction patterns of the photo-induced and the initial LT phases as a function of the time delay between the optical excitation and electron probe pulses. **b**, Difference between the diffraction patterns of the HT and LT phases (shown in Fig. 1). **c**, Relative intensity changes for a few selected reflections. On the left side of the vertical

green line, fast dynamical processes are shown; on the right side, slow dynamical processes are shown. The excitation fluence was 0.55 mJ cm^{-2} . The sample was photo-excited by 60-fs, 800-nm optical pulses at a repetition rate of 10 Hz. Error bars represent the standard deviation of measured values within the set of ten identical trials.

structural changes (see Supplementary Figs 9, 12 and 15), the observed experimental structure factor amplitudes can be written³⁰ as $|F_{\text{exp}}(k, t)| = |\eta_{\text{exc}} F_{\text{exc}}(k, t) + (1 - \eta_{\text{exc}}) F_{\text{LT}}(k)|$, where $F_{\text{exc}}(k, t)$ is the relevant part showing time-dependent changes and $F_{\text{LT}}(k)$ are the structure factor amplitudes of the initial LT phase. The structure factor amplitudes are weighted by the fraction of unit cells being altered, and remaining in the LT phase, respectively. Our model refinement calculations compare $|F_{\text{exp}}(k, t)|$ with simulated amplitudes $|F_{\text{sim}}(k, \xi)| = |\eta_{\text{exc}} F_{\text{sim, exc}}(k, \xi) + (1 - \eta_{\text{exc}}) F_{\text{sim, LT}}(k)|$ and maximize the Pearson correlation coefficient, $\gamma(|F_{\text{exp}}|, |F_{\text{sim}}|)$. This figure of merit γ varies between 0 (no correlation) and 1 (full correlation) and was found to provide well-behaved maxima across the full time series (see Supplementary Figs 14 and 15).

The term ξ is an atomic structural parameter composed of the three time-dependent generalized reaction coordinates (ξ_F, ξ_B, ξ_P) assigned to the motion of three independent dynamic groups: flat EDO-TTF molecules, bent EDO-TTF moieties and PF_6 anions, respectively. The configuration space was created from a linear interpolation (for each dynamic group) between the initial $\{\xi = (0, 0, 0)\}$ LT and the final

$\{\xi = (1, 1, 1)\}$ HT-type structures (see Supplementary Table 2). We believe that the reduction of the problem to such reaction coordinates (see Fig. 3a) should capture most of the physics and, at the same time, avoid unphysical results as a consequence of over-fitting. A value of $\eta_{\text{exc}} \approx 0.1$ was obtained by maximizing $\gamma(\eta_{\text{exc}})$ in the time interval of 0–10 ps (see Supplementary Fig. 15) and found to be in very good agreement with the estimated number of converted unit cells into the HT-type state obtained from equation (5) of the Supplementary Information and from the decrease at $t = 1$ ns in the structure factor amplitude averaged over $(h, k = 2n + 1, l)$ -reflections (where n is an integer) shown in Supplementary Fig. 9. Such reflections disappear in the HT phase owing to its higher symmetry (see Supplementary Information section 6.2).

These last two estimations of η_{exc} rely on the fact that by $t = 1$ ns the system has fully thermalized in a HT-type state and little residual thermal stress is left. These are fair assumptions considering that for $t > 500$ ps, the oscillations are damped, the estimated temperature increase is less than 50 K and the effect of residual stress is minimized for reflections satisfying Bragg's law²⁹. From optical reflectivity and

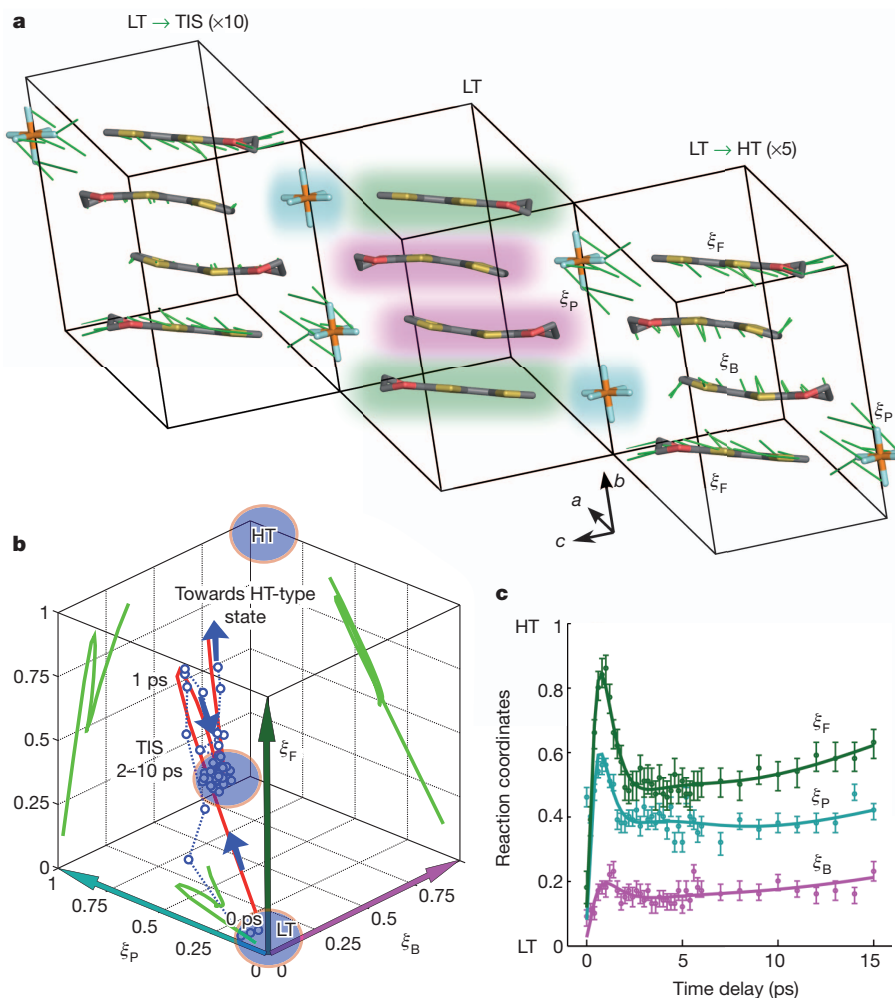


Figure 3 | Reaction coordinates and their temporal evolution in the formation of TIS. **a**, Three unit cells illustrating the arrangement of EDO-TTF molecules and PF_6 counter-ions in the LT phase. Selected dynamical groups are illustrated using different background colours in the central unit cell: flat EDO-TTF moieties have a green background, bent EDO-TTF moieties have a pink background, and PF_6 anions have a cyan background. Green sticks denote the atomic displacements from the initial LT structure to the TIS (unit cell on the left) and the HT-type structure (unit cell on the right); displacements were magnified 10 and 5 times, respectively, for clarity. The root-mean-square displacement between the LT structure and the TIS is 0.18 Å and between the LT and HT-type structures is 0.44 Å. The atomic coordinates of the TIS are

listed in Supplementary Table 2. **b**, Temporal evolution of ξ in a (ξ_F, ξ_B, ξ_P) -Cartesian configuration space. Circles indicate optimized structures. The red trace is a multi-exponential fit of the optimum correlation trajectory. The green trajectories are two-dimensional projections. **c**, Temporal evolution of each reaction coordinate and its contribution in the formation of the TIS. Error bars represent the full-width at half-maximum (FWHM) of the global maximum peaks of the Pearson correlation coefficient in reaction coordinate space. Solid traces are multi-exponential fits; see equation (9) in the Supplementary Information and Supplementary Table 1. See Supplementary Videos for full molecular movies highlighting the key molecular motions leading to the formation of the TIS state.

in situ transmission measurements we estimated that, for the excitation level used in the experiments, about 8% of the LT unit cells are initially photo-excited (see Supplementary Information section 2). The above observations indicate that the initial excitonic correlation length only extends over about one LT unit cell, promoting local structural changes.

Figure 3a depicts the selected dynamical groups (central unit cell) and the atomic displacements connecting the LT structure with the TIS (on the left) and the HT-type structures (on the right), respectively. The root-mean-square displacement between the LT structure and the TIS is 0.18 Å and between the LT and HT structures is 0.44 Å. Figure 3b summarizes the time-dependent changes of ξ , illustrated in a (ξ_F, ξ_B, ξ_P)-Cartesian configuration space. As can be seen from the two-dimensional projections, the motions of flat EDO-TTF moieties and of PF₆ counter-ions are correlated. The temporal evolution of each reaction coordinate is also shown in Fig. 3c, for clearness. The relative values among reaction coordinates suggest that optical excitation mostly drives the formation of the TIS through the sliding of flat EDO-TTF moieties and the correlated motion of PF₆ counter-ions. This observation is also supported by their respective overshoot values, $\Delta \equiv \xi(1 \text{ ps}) - \xi(5 \text{ ps})$, which are indicative of an over-damped half-cycle motion. The close proximity of PF₆ anions to EDO-TTF molecules and the observed correlation suggest that steric effects—that is, short-range repulsive interactions—are also important in the crystalline order and positioning of the anions.

The addition of nearest-neighbour Coulomb repulsion and electron-phonon interactions that modulate site energies through anion displacements were found to be crucial to reproduce the spectral signature of the ($D^{+1}D^{+0}D^{+1}D^{+0}$) excited state via an extended Peierls–Holstein–Hubbard model²³. Our results correspondingly indicate the motion of PF₆ counter-ions as a relevant component of the structural evolution of the excited electronic state and the formation of the TIS. Their connection with charge fluctuations suggests that PF₆ counter-ions need to be modified in order to gain control over electronic properties in (EDO-TTF)₂PF₆, akin to the anion replacement that confers superconductive properties at ambient pressure when going from (TMTSF)₂PF₆ to (TMTSF)₂ClO₄ (ref. 25). In contrast, the small relative amplitude for the unbending of EDO-TTF molecules in the formation of the TIS questions the relevance of this simple bending-type distortion mode in the formation of its charge disproportionates. Future FED experiments will focus on different sample orientations with improved spatial resolution and apply more general refinement algorithms for real-space reconstruction and *ab initio* electronic calculations to link dynamical structures with time-evolving charge distributions.

Received 10 August 2012; accepted 20 February 2013.

- Sciaini, G. & Miller, R. J. D. Femtosecond electron diffraction: heralding the era of atomically-resolved dynamics. *Rep. Prog. Phys.* **74**, 096101 (2011).
- Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Miller, R. J. D. An atomic-level view of melting using femtosecond electron diffraction. *Science* **302**, 1382–1385 (2003).
- Chergui, M. & Zewail, A. H. Electron and X-ray methods of ultrafast structural dynamics: advances and applications. *Phys. Chem. Chem. Phys.* **10**, 28–43 (2009).
- Baum, P. & Zewail, A. H. Breaking resolution limits in ultrafast electron diffraction and microscopy. *Proc. Natl Acad. Sci. USA* **103**, 16105–16110 (2006).
- Raman, R. K. *et al.* Direct observation of optically induced transient structures in graphite using ultrafast electron crystallography. *Phys. Rev. Lett.* **101**, 077401 (2008).
- Eichberger, M. *et al.* Snapshots of cooperative atomic motions in the optical suppression of charge density waves. *Nature* **468**, 799–802 (2010).
- Van Oudheusden, T. *et al.* Compression of subrelativistic space-charge-dominated electron bunches for single-shot femtosecond electron diffraction. *Phys. Rev. Lett.* **105**, 264801 (2010).
- Gao, M. *et al.* Full characterization of RF compressed femtosecond electron pulses using ponderomotive scattering. *Opt. Express* **20**, 12048–12058 (2012).
- Chatelain, R. P., Morrison, V. R., Godbout, C. & Siwick, B. J. Ultrafast electron diffraction with radio-frequency compressed electron pulses. *Appl. Phys. Lett.* **101**, 081901 (2012).

- Hastings, J. B. *et al.* Ultrafast time-resolved electron diffraction with megavolt electron beams. *Appl. Phys. Lett.* **89**, 184109 (2006).
- Musumeci, P., Moody, J. T., Scooby, C. M., Gutierrez, M. S. & Westfall, M. Laser-induced melting of a single crystal gold sample by time-resolved ultrafast relativistic electron diffraction. *Appl. Phys. Lett.* **97**, 063502 (2010).
- Tokita, S. *et al.* Single-shot femtosecond electron diffraction with laser-accelerated electrons: experimental demonstration of electron pulse compression. *Phys. Rev. Lett.* **105**, 215004 (2010).
- Murooka, Y. *et al.* Transmission-electron diffraction by MeV electron pulses. *Appl. Phys. Lett.* **98**, 251903 (2011).
- Kassier, G. H. *et al.* Photo-triggered pulsed cavity compressor for bright electron bunches in ultrafast electron diffraction. *Appl. Phys. B* **109**, 249–257 (2012).
- Rousse, A. *et al.* Non-thermal melting in semiconductors measured at femtosecond resolution. *Nature* **410**, 65–68 (2001).
- Sokolowski-Tinten, K. *et al.* Femtosecond X-ray measurement of coherent lattice vibrations near the Lindemann stability limit. *Nature* **422**, 287–289 (2003).
- Zamponi, F., Rothhardt, P., Stingl, J., Woerner, M. & Elsaesser, T. Ultrafast large-amplitude relocation of electronic charge in ionic crystals. *Proc. Natl Acad. Sci. USA* **109**, 5207–5212 (2012).
- Fritz, D. M. *et al.* Ultrafast bond softening in bismuth: mapping a solid's interatomic potential with X-rays. *Science* **315**, 633–636 (2007).
- Beaud, P. *et al.* Spatiotemporal stability of a femtosecond hard-X-ray undulator source studied by control of coherent optical phonons. *Phys. Rev. Lett.* **99**, 174801 (2007).
- Poulin, P. R. & Nelson, K. A. Irreversible organic crystalline chemistry monitored in real time. *Science* **313**, 1756–1760 (2006).
- Ota, A., Yamochi, H. & Saito, G. A novel metal-insulator phase transition observed in (EDO-TTF)₂PF₆. *J. Mater. Chem.* **12**, 2600–2602 (2002).
- Chollet, M. *et al.* Gigantic photoresponse in 1/4-filled-band organic salt (EDO-TTF)₂PF₆. *Science* **307**, 86–89 (2005).
- Onda, K. *et al.* Photoinduced change in the charge order pattern in the quarter-filled organic conductor (EDO-TTF)₂PF₆ with a strong electron-phonon interaction. *Phys. Rev. Lett.* **101**, 067403 (2008).
- Fukazawa, N. *et al.* Charge and structural dynamics in photoinduced phase transition of (EDO-TTF)₂PF₆ examined by picosecond time-resolved vibrational spectroscopy. *J. Phys. Chem. C* **116**, 5892–5899 (2012).
- Jérôme, D. The physics of organic superconductors. *Science* **252**, 1509–1514 (1991).
- Yonemitsu, K. & Maeshima, N. Photoinduced melting of charge order in a quarter-filled electron system coupled with different types of phonons. *Phys. Rev. B* **76**, 075105 (2007).
- Filatov, M. Antiferromagnetic interactions in the quarter-filled organic conductor (EDO-TTF)₂PF₆. *Phys. Chem. Chem. Phys.* **13**, 12328–12334 (2011).
- Iwano, K. & Shimoi, Y. Large electric-potential bias in an EDO-TTF tetramer as a major mechanism of charge ordering observed in its PF₆ salt: a density functional theory study. *Phys. Rev. B* **77**, 075120 (2008).
- Harb, M. *et al.* Excitation of longitudinal and transverse coherent acoustic phonons in nanometer free-standing films of (001) Si. *Phys. Rev. B* **79**, 094301 (2009).
- Coppens, P., Vorontsov, I. I., Graber, T., Gembicki, M. & Kovalevsky, A. Y. The structure of short-lived excited states of molecular complexes by time-resolved X-ray diffraction. *Acta Crystallogr. A* **61**, 162–172 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the contributions of J. Stampe, M. de Jong, M. Harb and S. G. Kruglik in developing the radio-frequency cavity system and of E. Pelletier in developing the laser system. We also thank K. Iwano and M. Hoshino and K. Yonemitsu for discussions. Funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada and the Canada Foundation for Innovation. This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (grant number 20110006) and the Global Centre of Excellence (G-COE) programme for Chemistry from The Ministry of Education, Culture, Sports, Science and Technology in Japan and by Creative Scientific Research (grant number 18GS0208) from The Japan Society for the Promotion of Science.

Author Contributions FED experiments were performed at University of Toronto in R.J.D.M.'s group. G. Sciaini initiated the EDO studies and conducted the work with R.J.D.M., C.L., G.M. and G. Sciaini performed the initial FED experiments at 1 kHz and 100 Hz. M.G., H.J.-R. and R.R.C. performed the final FED experiments with ultrabright electron pulses. M.G., H.J.-R. and A.M. did the data analysis. L.C.L. did the model structural refinement calculations and the Supplementary Video. M.G., H.J.-R. and R.R.C. performed the optical transmission measurements. K.O. and S.-y.K. performed the optical reflectivity measurements. Y.N., X.S., T.H., G. Saito and H.Y. provided the single crystals. C.L. was responsible for sample preparation for FED. M.G., L.C.L. and A.M. wrote the Supplementary Information. G. Sciaini, C.L., M.G. and R.J.D.M. wrote the manuscript with discussions among all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.D.M. (dwayne.miller@mpsd.cfel.de).

Terrestrial water fluxes dominated by transpiration

Scott Jasechko¹, Zachary D. Sharp¹, John J. Gibson^{2,3}, S. Jean Birks^{2,4}, Yi Yi^{2,3} & Peter J. Fawcett¹

Renewable fresh water over continents has input from precipitation and losses to the atmosphere through evaporation and transpiration. Global-scale estimates of transpiration from climate models are poorly constrained owing to large uncertainties in stomatal conductance and the lack of catchment-scale measurements required for model calibration, resulting in a range of predictions spanning 20 to 65 per cent of total terrestrial evapotranspiration (14,000 to 41,000 km³ per year) (refs 1–5). Here we use the distinct isotope effects of transpiration and evaporation to show that transpiration is by far the largest water flux from Earth's continents, representing 80 to 90 per cent of terrestrial evapotranspiration. On the basis of our analysis of a global data set of large lakes and rivers, we conclude that transpiration recycles 62,000 ± 8,000 km³ of water per year to the atmosphere, using half of all solar energy absorbed by land surfaces in the process. We also calculate CO₂ uptake by terrestrial vegetation by connecting transpiration losses to carbon assimilation using water-use efficiency ratios of plants, and show the global gross primary productivity to be 129 ± 32 gigatonnes of carbon per year, which agrees, within the uncertainty, with previous estimates⁶. The dominance of transpiration water fluxes in continental evapotranspiration suggests that, from the point of view of water resource forecasting, climate model development should prioritize improvements in simulations of biological fluxes rather than physical (evaporation) fluxes.

Unlike river discharges to the oceans⁷, the global fluxes of evaporation and transpiration are poorly constrained owing to a lack of methodology to decouple these two water fluxes at the catchment scale. Stable isotope ratios of oxygen (¹⁸O/¹⁶O) and hydrogen (²H/¹H) in water can be used to separate transpiration from evaporation⁸, because the two processes have different effects on these ratios in water. The physical process of evaporation enriches residual water in the heavy isotopes of oxygen and hydrogen, whereas the biological process of transpiration does not produce an isotopic fractionation, assuming an isotopic steady state over annual timescales^{8–11}. The pathway water takes after falling as precipitation within a catchment includes mixing, evaporation (fractionation labelled) and transpiration (non-fractionation labelled), until the remaining water accumulates in a downstream lake or river. Each of these catchment processes is ultimately recorded by the isotopic composition of the lake's water. We have compiled a data set of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values of large lake waters and capitalize on dissimilar isotope effects between evaporation and transpiration to decouple and quantify these two freshwater losses from Earth's surface (isotope content is given by $(R_{\text{sample}}/R_{\text{V-SMOW}} - 1) \times 10^3 \text{‰}$, where R is ¹⁸O/¹⁶O for $\delta^{18}\text{O}$ and ²H/¹H for $\delta^2\text{H}$, and V-SMOW represents standard mean ocean water).

To proceed with this calculation, we first report on the stable oxygen and hydrogen isotope compositions of Earth's large lakes (Fig. 1). The isotopic compositions of lake waters show a broad range in $\delta^{18}\text{O}$ and

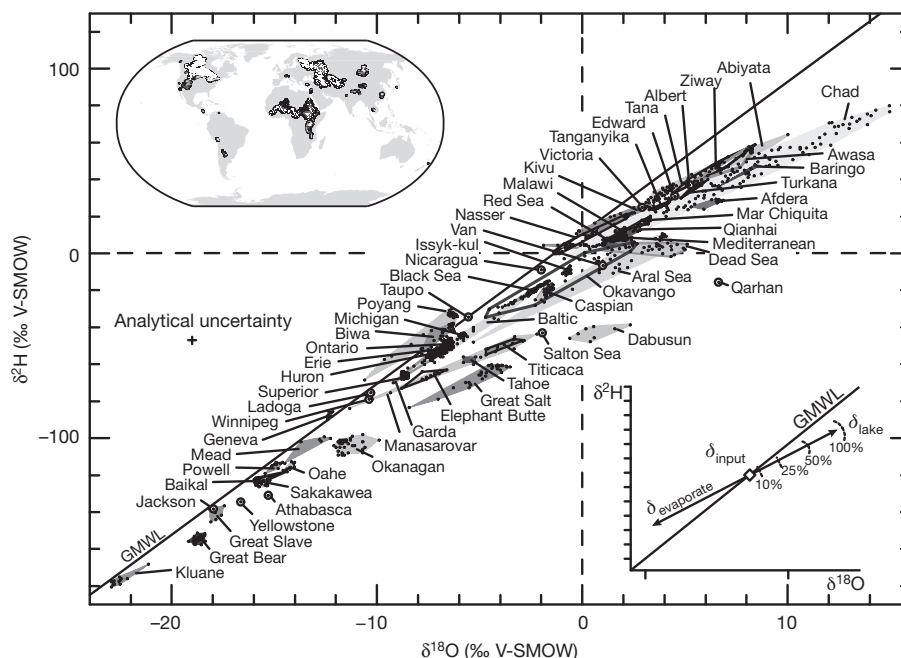


Figure 1 | $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values of large lakes and semi-enclosed seas. The global meteoric water line¹² (GMWL) is shown. The map at top left shows catchment areas covered by the data set. The schematic graph at bottom right

shows water inputs to a lake (diamond) and the evaporation trajectory of a lake (percentages refer to evaporation amount).

¹Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA. ²Alberta Innovates – Technology Futures, Vancouver Island Technology Park, Victoria, British Columbia V8Z 7X8, Canada. ³Department of Geography, University of Victoria, Victoria, British Columbia V8W 3R4, Canada. ⁴Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

$\delta^2\text{H}$ values: -23‰ to $+15\text{‰}$ and -180‰ to $+80\text{‰}$, respectively. Well-mixed lakes (for example, the North American Great Lakes and Lake Baikal) have relatively homogenous stable isotope compositions, whereas perennially stratified or shallow lakes tend to have greater isotopic variability (for example Lake Kivu and Lake Chad). Headwater lakes located at high latitudes and altitudes (for example, Kluane Lake) have the lowest $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values, whereas the closed basin lakes of eastern Africa have the highest $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values (for example, Lake Afdera and Lake Turkana). Global lakes do not follow one systematic evaporation trend, reflecting the unique climatology and hydrology of each individual lake catchment. The global meteoric water line plotted in Fig. 1 is a regression of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values of precipitation samples on a global scale¹². This regression produces a $\delta^2\text{H}/\delta^{18}\text{O}$ slope of eight that can be closely reconciled by liquid–vapour isotope effects at chemical equilibrium¹³. However, the disequilibrium process of evaporation results in a strong kinetic isotope effect, with the light isotopologues preferentially partitioned into the vapour phase. This results in $\delta^2\text{H}/\delta^{18}\text{O}$ slopes of less than eight, driving the isotope composition of lake waters ‘below’ the global meteoric water line. Information on the percentage of evaporative losses is retained by the difference between the lake’s isotope composition and that of waters entering a basin (δ_{input} ; Fig. 1, inset schematic graph). Our global data compilation shows that nearly all lakes fall to the right of the global meteoric water line in $\delta^{18}\text{O}$ – $\delta^2\text{H}$ space as a result of evaporation, except in special cases where waters evaporate upwind and re-precipitate in a downwind lake basin (for example Lake Biwa). In what follows, we develop equations describing a stable isotope mass balance of waters within a lake catchment to estimate the percentage of catchment transpiration, and apply these equations to $\delta^{18}\text{O}$ and $\delta^2\text{H}$ data for large lake waters.

A lake catchment in hydrologic steady state can be described by a balance between water inputs (I , precipitation and inflows from upstream lakes) and water losses such as precipitation intercepted by vegetation (xP , where x is the fraction of intercepted precipitation for the catchment), open-water and soil evaporation (E), transpiration (T) and liquid losses to rivers or groundwater discharge (Q):

$$I = xP + E + T + Q \quad (1)$$

Similarly, a stable isotope mass balance of a lake catchment is obtained by considering the isotopic composition of each water flux (δ_i , δ_E and so on):

$$\delta_i I = \delta_P xP + \delta_E E + \delta_T T + \delta_Q Q \quad (2)$$

By combining equations (1) and (2), we develop a new equation describing transpiration losses from a catchment:

$$T = \frac{I(\delta_i - \delta_E) - Q(\delta_Q - \delta_E) - xP(\delta_P - \delta_E)}{\delta_T - \delta_E} \quad (3)$$

Each parameter in equation (3) can be estimated from gridded data sets or lake-specific studies, except for the isotope composition of evaporate (δ_E). To calculate δ_E (ref. 14), we use an evaporation model based on laboratory-derived liquid–vapour fractionation factors¹³. The isotope composition of soil and open-water evaporate are grouped into one term (δ_E), and the isotopic composition of transpired moisture (δ_T) is calculated using both shallow and deep-water sources under the assumption that the catchment is in steady state at an annual time step¹¹ (on average, water molecules spend multiple years within lakes examined here). Physical, isotopic and hydroclimatic data sets for each lake are compiled from available reanalysis and interpolated data^{15–18} and are used in equation (3) to calculate catchment transpiration (Methods).

We find that transpiration accounts for more than two-thirds of total surface water evapotranspiration for 85% of the catchments examined (Fig. 2 and Fig. 3a). Remarkably, transpiration also accounts for the majority of evapotranspiration in desert catchments (average, 75%; range, 35% to 95%). *In situ* transpiration measurements in

deserts range from 7% to 80% of evapotranspiration¹⁹, in large part because precipitation rates are highest in the headwaters of desert catchments, thereby increasing the importance of these forested ecosystems to the catchment’s water balance. Transpiration rates range from less than 100 mm yr^{-1} to approximately $1,300 \text{ mm yr}^{-1}$ (Fig. 2 and Supplementary Information, section 1). Our results show that even though open-water evaporation may locally occur at higher rates than transpiration, the fraction of total evapotranspiration represented by evaporation is severely limited by the small areas of open water on Earth’s continents (approximately 3%, globally²⁰). Therefore, we posit that the biological pump of water into the atmosphere during photo-synthetic gas exchange (that is, transpiration), rather than the physical process of evaporation, dominates water losses from the continents. Because plant roots are able to tap into groundwater and soil-water reservoirs, transpiration effectively moves deep sources of water into the atmosphere, whereas evaporation is only effective for water at or near the surface, which explains the very high proportion of transpiration to the overall evapotranspiration flux.

Our results are supported by a cross-plot comparison of isotope-based and conventional open-water evaporation rates ($R^2 = 0.78$ (squared correlation coefficient), slope = 0.92; Supplementary Fig. 1); geographic similarity between compiled *in situ* transpiration measurements^{9,10} (of, for example, sap flow) at the forest stand level and our estimates (Fig. 3); and agreement between $^{18}\text{O}/^{16}\text{O}$ - and $^2\text{H}/^1\text{H}$ -based evaporation rates using equation (3) ($R^2 = 0.78$, slope = 0.94; Supplementary Figs 2 and 3 and Supplementary Information, section 1). We also note that the time step of our calculated transpiration fluxes ranges from 1 to 1,000 years, averaging 40 years, as dictated by the hydrologic residence time of each lake (Supplementary Table 3). To scale up our calculation to Earth’s ice-free land surface, and to provide a fourth check corroborating our catchment transpiration results, we estimate the global transpiration from Earth’s landmasses (excluding Antarctica) from a stable isotope mass balance of Earth’s entire freshwater reservoir. This estimate is based on the deuterium excess parameter, which includes information contained in both $^{18}\text{O}/^{16}\text{O}$ and $^2\text{H}/^1\text{H}$ ($d = \delta^2\text{H} - 8\delta^{18}\text{O}$). We

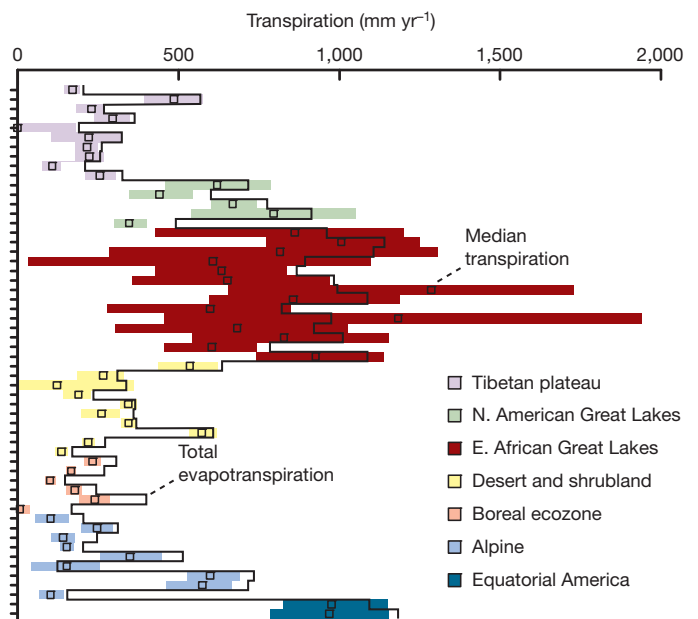


Figure 2 | Transpiration water losses for 56 lake catchments grouped by ecoregion ($^{18}\text{O}/^{16}\text{O}$ -based results). Each coloured bar represents results for a single lake catchment. Extents of bars show 25th and 75th percentiles of Monte Carlo simulations. Median transpiration (T ; square) outputs of Monte Carlo simulations and total evapotranspiration losses (solid line) are shown. The median result is close to the total evapotranspiration for most of the lakes, demonstrating the dominant role of transpiration in total evapotranspiration losses.

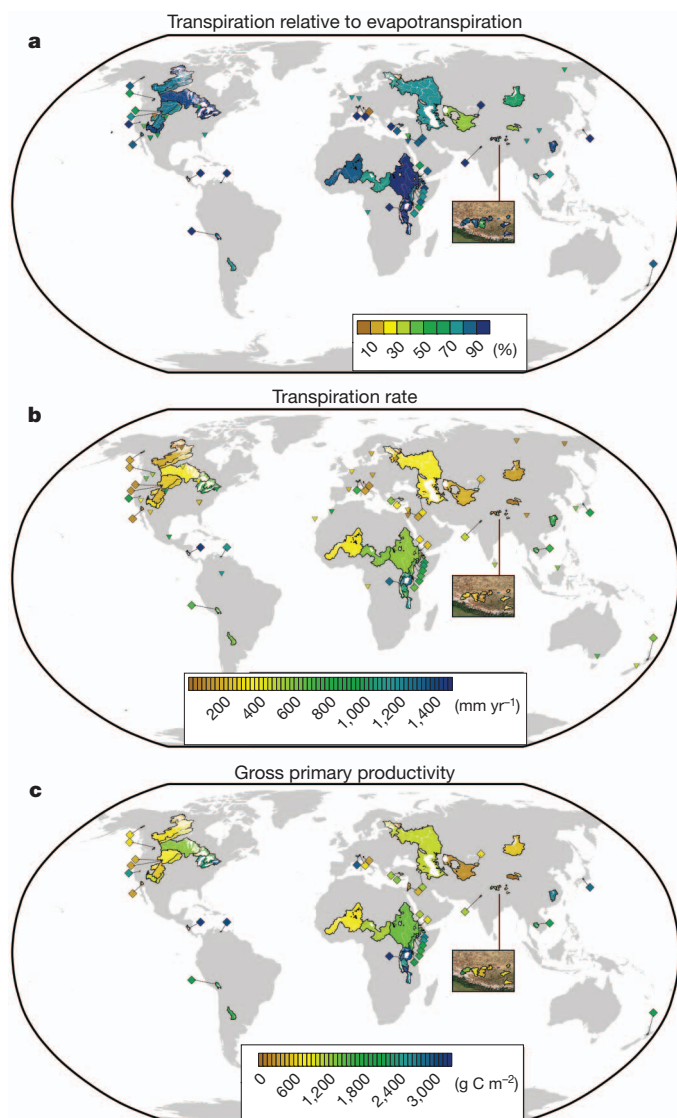


Figure 3 | Transpiration and carbon fluxes within 73 lake catchments. **a**, Transpiration losses as a percentage of total evapotranspiration. **b**, Transpiration rates. **c**, Gross primary productivity for 10% of Earth's continental area. Coloured diamonds are shown for small basins as a visual aid. Inverted triangles represent compiled *in situ* transpiration measurements (for example sap flow⁹).

obtain a similar expression to equation (3) based on the deuterium excess:

$$T = \frac{P(d_p - d_E) - Q(d_Q - d_E) - xP(d_p - d_E)}{d_T - d_E} \quad (4)$$

Terrestrial precipitation ($P = 110,000 \pm 10,000 \text{ km}^3 \text{ yr}^{-1}$ (ref. 17), $d_p = 9.5 \pm 1\text{‰}$ (ref. 12)) is the only input of water to the continents, and water is lost through river discharges to the oceans ($Q = 37,300 \pm 7,000 \text{ km}^3 \text{ yr}^{-1}$ (ref. 7), $d_Q = 6.8 \pm 3.8\text{‰}$; Supplementary Information, section 2), terrestrial evaporation ($d_E = 75 \pm 30\text{‰}$ (this work)), transpiration ($d_T = 8 \pm 3\text{‰}$) or interception by vegetation ($xP = 7,500 \pm 1,500 \text{ km}^3 \text{ yr}^{-1}$ (refs 16, 17), $d_p = 9.5 \pm 1\text{‰}$ (ref. 12)). Solving equation (4) shows that transpiration accounts for 80% to 90% of terrestrial evapotranspiration (respectively the 25th and 75th percentiles of a Monte Carlo sensitivity analysis). Volumetrically, transpiration converts $62,000 \pm 8,000 \text{ km}^3 \text{ yr}^{-1}$ of liquid water into atmospheric vapour, requiring $33 \pm 4 \text{ W m}^{-2}$ of latent heat, or roughly half of all solar energy absorbed by the continents²¹ (approximately 70 W m^{-2}). Results show

that 90% of precipitation falling on land¹⁸ ($111,000 \text{ km}^3 \text{ yr}^{-1}$) is already appropriated to ecosystems for either primary production ($62,000 \pm 8,000 \text{ km}^3 \text{ yr}^{-1}$) or as aquatic habitat in rivers⁷ ($37,000 \text{ km}^3 \text{ yr}^{-1}$), an important consideration for diversion of in-stream flows.

We can use our transpiration fluxes to calculate carbon assimilation by terrestrial vegetation by linking the water and carbon cycles²². The molar ratio of CO_2 assimilated during photosynthesis to transpired H_2O is known as water-use efficiency (WUE) and is dependent on a variety of factors including the type of photosynthetic pathway used by a particular plant species (C_3 , C_4 or CAM) and atmospheric conditions such as vapour pressure deficit and CO_2 concentration²³. We compile WUE data and couple these to atmospheric vapour pressure deficits and C_3 – C_4 vegetation abundances to develop global grids of WUE (Methods). Catchment transpiration fluxes and gridded WUEs are applied to calculate gross primary production within each catchment (Fig. 3c). On the global scale, we weight our global grids of WUE according to vegetation density and calculate the global WUE of the terrestrial biosphere to be $3.2 \pm 0.9 \text{ mmol CO}_2 \text{ per mol H}_2\text{O}$. Applying this ratio to our global transpiration flux ($62,000 \pm 8,000 \text{ km}^3 \text{ yr}^{-1}$ of H_2O), we calculate gross primary production to be $129 \pm 32 \text{ GtC yr}^{-1}$. This flux is consistent with a recent estimate of $123 \pm 8 \text{ GtC yr}^{-1}$ (ref. 6) and provides a fifth line of support for the large transpiration fluxes reported in our work.

Linkages made here between the water and carbon cycles highlight a new stable-isotope-based methodology that can be used to monitor and map ongoing changes to Earth's water cycle^{24,25} as well as modifications to carbon assimilation rates under increased atmospheric temperatures and CO_2 concentrations. Our results show that the water and carbon cycles are linked in such a way that transpiration must account for more than 80% of continental evapotranspiration to maintain a mass balance between these two biogeochemical fluxes (plant transpiration and CO_2 uptake). Given the importance of transpiration, it follows that the physiological response of vegetation to a warmer and CO_2 -enriched atmosphere will have a dominant effect on future changes to evapotranspiration and the terrestrial hydrological cycle. Furthermore, changes in natural ecosystems via land-use modifications or climate changes will have notable effects on river discharges and, consequently, fluvial sediment loads, chemical weathering on continents, and atmospheric latent heat transport.

Climate change is expected to affect global transpiration³. Considering the dominance of transpiration in continental evapotranspiration shown here, future changes in global transpiration will affect land temperatures by altering latent heat fluxes from continents, and will also change the fraction of precipitation entering rivers. Our catchment-scale results can be applied as a calibration tool for climate models, which should shift the prevailing focus on physical climate data towards ecosystem water requirements and so result in better predictions of continental evapotranspiration and water recycling in a warmer future climate.

METHODS SUMMARY

In equation (3), hydrologic inputs (I) include precipitation¹⁷ and upstream lake inflows. Catchment losses include interception (xP ; ref. 16), liquid outflows (Q ; lake-specific data), evaporation (E) and transpiration (T). The isotopic composition of precipitation (δ_p) is computed from monthly grids¹⁵ weighted spatially (grid cell, i) and temporally (month, j) to precipitation distribution as follows:

$$\delta_p = \frac{1}{\sum_{i=1}^n P_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} P_j \delta_{p_j}}{\sum_{j=1}^{12} P_j} \right) P_i \quad (5)$$

For chain lakes, inflows from an upstream lake are included in the calculation of δ_i . For lakes not in equilibrium with current climate, δ_i is calculated from the intercept between a computed evaporation trend²⁶ and the global meteoric water line¹². The isotopic composition of lake water is used to estimate that of liquid outflows (that is, $\delta_Q = \delta_{\text{lake}}$). The isotopic composition of transpired moisture (δ_T) is estimated by weighting the isotopic composition of precipitation¹⁵ spatially (i) to

mean long-term normalized difference vegetation indices (NDVI; proxy for vegetation density). A range of two temporal (j) weighting approaches¹⁰ is used for δ_T , one weighted to growing season (NDVI values (equation (5)) and another to monthly precipitation (equation (6)):

$$\delta_{T-\text{SHALLOW}} = \frac{1}{\sum_{i=1}^n \text{NDVI}_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} \text{NDVI}_j \delta_{P_j}}{\sum_{j=1}^{12} \text{NDVI}_j} \right) \text{NDVI}_i \quad (6)$$

$$\delta_{T-\text{DEEP}} = \frac{1}{\sum_{i=1}^n \text{NDVI}_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} P_j \delta_{P_j}}{\sum_{j=1}^{12} P_j} \right) \text{NDVI}_i \quad (7)$$

Finally, the isotopic composition of evaporate is computed using an evaporation model¹⁴:

$$\delta_E = \frac{(\delta_{\text{lake}} - \varepsilon^*) / \alpha^* - h \delta_A - \varepsilon_K}{1 - h + \varepsilon_K} \quad (8)$$

This takes into account temperature-dependent¹⁷ equilibrium¹³ (α^* ; $\varepsilon^* = \alpha^* - 1$) and humidity-dependent¹⁷ (h) kinetic (ε_K ; ref. 27) fractionation factors normalized to lake temperatures²⁸. The isotopic composition of atmospheric vapour (δ_A) is computed on the basis of a precipitation-equilibrium assumption^{13,15,17,26} and by using isotope-enabled climate model grids¹⁸. Humidity, temperature and δ_A are each weighted to the monthly evaporation amount. A Monte Carlo simulation is used to assess calculation uncertainty embedded in each input parameter (Supplementary Information, sections 3 and 4). Plant water-use efficiency is calculated by applying growing-season daytime vapour pressure deficit^{17,29} (VPD) to C_3 ($\text{WUE}_{C_3} = 4.21(\text{VPD})^{-0.67}$ mmol CO_2 per mol H_2O) and C_4 ($\text{WUE}_{C_4} = 6.91(\text{VPD})^{-0.40}$ mmol CO_2 per mol H_2O) vegetation abundances³⁰.

Full Methods and any associated references are available in the online version of the paper.

Received 6 September 2012; accepted 4 February 2013.

Published online 3 April 2013.

- Lawrence, D. M., Thornton, P. E., Oleson, K. W. & Bonan, G. B. Partitioning of evaporation into transpiration, soil evaporation, and canopy evaporation in a GCM: impacts on land-atmosphere interaction. *J. Hydrometeorol.* **8**, 862–880 (2007).
- Alton, P., Fisher, R., Los, S. & Williams, M. Simulations of global evapotranspiration using semiempirical and mechanistic schemes of plant hydrology. *Glob. Biogeochem. Cycles* **23**, GB4023 (2009).
- Cao, L., Bala, G., Caldeira, K., Nemani, R. & Ban-Weiss, G. Importance of carbon dioxide physiological forcing to future climate change. *Proc. Natl Acad. Sci. USA* **107**, 9513–9518 (2010).
- Ito, A. & Motoko, I. Water-use efficiency of the terrestrial biosphere: a model analysis focusing on interactions between the global carbon and water cycles. *J. Hydrometeorol.* **13**, 681–694 (2012).
- Gerten, D. *et al.* Contemporary “green” water flows: simulations with a dynamic global vegetation and water balance model. *Phys. Chem. Earth* **30**, 334–338 (2005).
- Beer, C. *et al.* Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329**, 834–838 (2010).
- Dai, A. & Trenberth, K. E. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *J. Hydrometeorol.* **3**, 660–687 (2002).
- Yakir, D. & Wang, X. F. Fluxes of CO_2 and water between terrestrial vegetation and the atmosphere estimated from isotope measurements. *Nature* **380**, 515–517 (1996).
- Williams, D. *et al.* Evapotranspiration components determined by stable isotope, sap flow and eddy covariance techniques. *Agric. For. Meteorol.* **125**, 241–258 (2004).
- Dawson, T. E. Determining water use by trees and forests from isotopic, energy balance and transpiration analyses: the roles of tree size and hydraulic lift. *Tree Physiol.* **16**, 263–272 (1996).
- Welp, L. R. *et al.* $\delta^{18}\text{O}$ of water vapor, evapotranspiration and the sites of leaf water evaporation in a soybean canopy. *Plant Cell Environ.* **31**, 1214–1228 (2008).
- Rozanski, K., Araguas-Araguas, L. & Gonfiantini, R. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K. *et al.*) 1–36 (Am. Geophys. Union, 1993).
- Horita, J. & Wesolowski, D. Liquid-vapour fractionation of oxygen and hydrogen isotopes of water from the freezing to the critical temperature. *Geochim. Cosmochim. Acta* **58**, 3425–3437 (1994).
- Craig, H. & Gordon, L. I. in *Stable Isotopes in Oceanographic Studies and Paleotemperatures* (ed. Tongiorgi, E.) 9–130 (Lab. Geol. Nucl., 1965).
- Bowen, G. J. & Revenaugh, J. Interpolating the isotopic composition of modern meteoric precipitation. *Wat. Resour. Res.* **39**, 1299 (2003).
- Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M. & Dolman, A. J. Global canopy interception from satellite observations. *J. Geophys. Res.* **115**, D16122 (2010).
- New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25 (2002).
- Yoshimura, K., Kanamitsu, M., Noone, D. & Oki, T. Historical isotope simulation using reanalysis atmospheric data. *J. Geophys. Res.* **113**, D19108 (2008).
- Reynolds, J. F., Kemp, P. R. & Tenhunen, J. D. Effects of long-term rainfall variability on evapotranspiration and soil water distribution in the Chihuahuan desert: a modeling analysis. *Plant Ecol.* **150**, 145–159 (2000).
- Downing, J. A. *et al.* The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **51**, 2388–2397 (2006).
- Trenberth, K. E., Fasullo, J. T. & Kiehl, J. Earth's global energy budget. *Bull. Am. Meteorol. Soc.* **90**, 311–323 (2009).
- Beer, C., Reichstein, M., Ciais, P., Farquhar, G. D. & Papale, D. Mean annual GPP of Europe derived from its water balance. *Geophys. Res. Lett.* **34**, L05401 (2007).
- Farquhar, G. D., Ehleringer, J. R. & Hubick, K. T. Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 503–537 (1989).
- Jung, M. *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954 (2010).
- Durack, P. J., Wijffels, S. E. & Matear, R. J. Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science* **336**, 455–458 (2012).
- Gibson, J. J., Birks, S. J. & Edwards, T. W. D. Global prediction of δ_A and $\delta^2\text{H}$ - $\delta^{18}\text{O}$ evaporation slopes for lakes and soil water accounting for seasonality. *Glob. Biogeochem. Cycles* **22**, GB2031 (2008).
- Gonfiantini, R. in *Handbook of Environmental Isotope Geochemistry* Vol. 2: *The Terrestrial Environment* (eds Fritz, P. & Fontes, J.-Ch.) 113–163 (Elsevier, 1986).
- Buck, A. L. New equations for computing vapour pressure and enhancement factor. *J. Appl. Meteorol.* **20**, 1527–1532 (1981).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Global Energy and Water Cycle Experiment. International Satellite Land-Surface Climatology Project. <http://www.gewex.org/islscpdata.htm> (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. W. D. Edwards, T. Gleeson and M. C. Molles Jr for comments on the manuscript, and are grateful to O. Kwiczen, D. G. Miralles, B. K. Nyarko, K. Yoshimura and F. Yuan for providing access to isotope and gridded data sets. Support for this work was provided by a graduate fellowship awarded to S.J. by the Caswell Silver Foundation through the University of New Mexico.

Author Contributions S.J. designed the study, compiled each data set, did the geographic information system and remote sensing work, developed the equations, did the water balance and carbon flux calculations, and wrote the paper. Z.D.S., J.J.G., S.J.B., Y.Y. and P.J.F. discussed the results, commented on the manuscript and contributed to text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J. (jasechko@unm.edu).

METHODS

For each basin, eight input terms are required for calculation of transpiration losses using equation (3): I , Q , x , δ_Q , δ_I , δ_T , δ_P and δ_E . Water inputs (I) include precipitation and upstream lake inflows. Precipitation inputs are obtained from high-resolution physical climate grids¹⁷. Upstream chain lake inflows are retrieved from lake-specific sources, as are liquid outflows (Q) for each lake. The proportion of incident precipitation that is intercepted and returned to the atmosphere (x) is obtained from satellite-based gridded data¹⁶. The isotope composition of liquid outflows (δ_Q) from each lake is obtained from epilimnion $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values for samples nearest to a lake's outflow ($\delta_{\text{lake}} = \delta_Q$). The isotope composition of precipitation entering a basin is calculated by weighting spatially (grid cell, i) and temporally (month, j) to monthly precipitation amount (P) as follows:

$$\delta_P = \frac{1}{\sum_{i=1}^n P_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} P_j \delta_{P_j}}{\sum_{j=1}^{12} P_j} \right)_i P_i \quad (5)$$

Monthly δ_P estimates¹⁵ and monthly precipitation amounts are obtained from gridded data sets. For headwater lakes, precipitation is the sole input (that is, $I = P$ and $\delta_I = \delta_P$); for chain lakes, δ_P is flux-weighted against the isotopic composition of riverine inputs entering the catchment from upstream chain lakes. Large lakes with residence times longer than approximately 300 years (for example, Lake Baikal) are in an isotopic disequilibrium with current climate. For these lakes, a $\delta^2\text{H}/\delta^{18}\text{O}$ evaporation slope is calculated²⁶ using the lake isotope data. The intercept of the resulting 'evaporation line' with the global meteoric water line¹² is applied as a mean 'long-term' estimate of δ_I . This approach is also applied to headwater lakes where grids produce unrealistic δ_I estimates ($\delta_I > \delta_{\text{lake}}$). Lakes with an outflow that periodically reverses flow (Tonlé Sap and Poyang Lake) or mixes with a geographically separate lake (Lake Michigan and Lake Huron) are treated specially, and the isotopic composition and flux of return flows are included in the computation of δ_I .

The isotopic composition of transpired moisture is calculated using an average of two approaches (both based on similar δ_P grids¹⁵), the first representing the isotope composition of shallow waters during the growing season and the second estimating the annual isotope composition of recharge. Plants tapping shallow water sources draw on precipitation falling during the growing season, whereas deep-rooted vegetation transpires ground waters that more closely represent the recharge-weighted isotope composition of precipitation¹⁰.

To estimate the isotope composition of transpired moisture (δ_T) for shallow 'growing season' waters, the isotope composition of precipitation (δ_P) is weighted to a proxy for chlorophyll abundance using long-term monthly mean values of normalized difference vegetation indices (NDVI; equation (5)). Monthly NDVI values less than zero were assigned a value of zero, because these cells host minor amounts of photosynthetic activity:

$$\delta_{T-\text{SHALLOW}} = \frac{1}{\sum_{i=1}^n \text{NDVI}_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} \text{NDVI}_j \delta_{P_j}}{\sum_{j=1}^{12} \text{NDVI}_j} \right)_i \text{NDVI}_i \quad (6)$$

To estimate the isotope composition of transpired moisture from vegetation with deep roots, we weight the isotope composition of precipitation temporally to monthly precipitation amount, and then consider the spatial distribution of vegetation by applying long-term annual average NDVI indices:

$$\delta_{T-\text{DEEP}} = \frac{1}{\sum_{i=1}^n \text{NDVI}_i} \sum_{i=1}^n \left(\frac{\sum_{j=1}^{12} P_j \delta_{P_j}}{\sum_{j=1}^{12} P_j} \right)_i \text{NDVI}_i \quad (7)$$

The range of values from the two approaches is used to estimate δ_T (see Supplementary Information, section 3 for treatment of uncertainty in each input parameter).

The isotope composition of evaporated moisture is estimated using an evaporation model developed to calculate the isotopic composition of evaporate¹⁴:

$$\delta_E = \frac{(\delta_{\text{lake}} - \varepsilon^*) / \alpha^* - h \delta_A - \varepsilon_K}{1 - h + \varepsilon_K} \quad (8)$$

where δ_{lake} is the isotope composition of each lake, α^* is the temperature-dependent¹⁷ equilibrium liquid-vapour fractionation factor¹³ (with $\varepsilon^* = \alpha^* - 1$), h is the relative humidity¹⁷ normalized to surface water temperatures and $\varepsilon_K = C_K(1 - h)$ is a kinetic separation factor with C_K representing resistance ratios for various stable water isotopologues ($C_K = 14.2$ for the $\delta^{18}\text{O}$ model and $C_K = 12.5$ for the $\delta^2\text{H}$ model²⁷).

To calculate δ_E , four inputs are required: atmospheric specific humidity, lake water temperature (T_L), air temperature (T_A) and the isotopic composition of atmospheric moisture (δ_A). Specific humidity values are calculated²⁸ using long-term monthly mean temperature and relative humidity grids¹⁷ (h_A), lake surface temperatures are used to calculate monthly saturation vapour pressures, and then relative humidity is computed using the specific humidity and computed saturation vapour pressure. The isotope composition of atmospheric water vapour (δ_A) is obtained in two ways. First, the isotope composition of atmospheric moisture can be estimated by assuming that the isotope composition of precipitation reflects the isotopic atmospheric vapour offset by equilibrium isotope effects²⁶ (calculated with monthly gridded δ_P values¹⁵ and air temperatures¹⁷, applied to liquid-vapour equilibrium fractionation factors¹³). Alternatively, values for δ_A are derived from long-term monthly average outputs from an isotope-enabled global climate model (IsoGSM¹⁸). An average δ_A value is taken from the two approaches and used as a first estimate (Supplementary Information, section 3). Finally, atmospheric vapour $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values, humidity, and lake and air temperatures are weighted temporally to monthly evaporation amount for each lake. This is a crucial step for lakes that experience large seasonal variations in evaporation fluxes. Uncertainty in each input parameter is assessed by a Monte Carlo analysis (Supplementary Information, sections 3 and 4). Plant-scale water-use efficiency grids are developed for Earth from a compilation of data sets (Supplementary Table 6 and Supplementary Figs 5 and 6). Daytime vapour pressure deficit (VPD) is calculated at a monthly time step using humidity¹⁷ and daytime temperature (maximum daily temperature minus average daily temperature²⁹), and is weighted to growing season via NDVI indices. These growing-season daytime vapour pressure deficits are input into VPD-WUE regressions from compiled data for C_3 ($\text{WUE}_{C_3} = 4.21(\text{VPD})^{-0.67}$ mmol CO_2 per mol H_2O) and C_4 ($\text{WUE}_{C_4} = 6.91(\text{VPD})^{-0.40}$ mmol CO_2 per mol H_2O) vegetation, and the respective proportions of C_3 and C_4 plants within each grid cell³⁰ are applied to develop WUE grids for Earth. WUE grids are applied to catchment transpiration fluxes to calculate gross primary production within each lake basin.

Earliest evidence for the use of pottery

O. E. Craig¹, H. Saul^{1,2}, A. Lucquin¹, Y. Nishida², K. Taché¹, L. Clarke^{3,4}, A. Thompson⁵, D. T. Alft¹, J. Uchiyama⁶, M. Ajimoto⁷, K. Gibbs⁸, S. Isaksson^{9,10}, C. P. Heron³ & P. Jordan¹¹

Pottery was a hunter-gatherer innovation that first emerged in East Asia between 20,000 and 12,000 calibrated years before present^{1,2} (cal BP), towards the end of the Late Pleistocene epoch, a period of time when humans were adjusting to changing climates and new environments. Ceramic container technologies were one of a range of late glacial adaptations that were pivotal to structuring subsequent cultural trajectories in different regions of the world, but the reasons for their emergence and widespread uptake are poorly understood. The first ceramic containers must have provided prehistoric hunter-gatherers with attractive new strategies for processing and consuming foodstuffs, but virtually nothing is known of how early pots were used. Here we report the chemical analysis of food residues associated with Late Pleistocene pottery, focusing on one of the best-studied prehistoric ceramic sequences in the world, the Japanese Jōmon. We demonstrate that lipids can be recovered reliably from charred surface deposits adhering to pottery dating from about 15,000 to 11,800 cal BP (the Incipient Jōmon period), the oldest pottery so far investigated, and that in most cases these organic compounds are unequivocally derived from processing freshwater and marine organisms. Stable isotope data support the lipid evidence and suggest that most of the 101 charred deposits analysed, from across the major islands of Japan, were derived from high-trophic-level aquatic food. Productive aquatic ecotones were heavily exploited by late glacial foragers³, perhaps providing an initial impetus for investment in ceramic container technology, and paving the way for further intensification of pottery use by hunter-gatherers in the early Holocene epoch. Now that we have shown that it is possible to analyse organic residues from some of the world's earliest ceramic vessels, the subsequent development of this critical technology can be clarified through further widespread testing of hunter-gatherer pottery from later periods.

On the basis of evidence from East Asia, North Africa, the Americas and Northern Europe, it is now widely accepted that pottery production precedes the emergence of farming and should be considered a hunter-gatherer innovation. From radiocarbon dates of an archaeological sequence excavated at Xianrendong Cave, Jiangxi Province, China, it has been proposed that ceramic container technology extends in age to about 20,000–19,000 cal BP¹, some 10,000 yr before the earliest evidence of rice cultivation in that region. Regional case studies^{4,5} have explored the origins, adoption and dispersal of ceramic container technology, dates are being refined and pushed back, and models are emerging that encompass economic, technological, social and environmental factors that both facilitated, and potentially inhibited, pottery innovation in different parts of the world⁶. Yet the actual uses of these early pottery vessels have not been addressed beyond assumed functional roles such as cooking and storage, or speculative associations with available animal and plant resources^{7–9}.

An understanding of how pottery was used is necessary to elucidate both the social and the economic context for its innovation and also to

explain why the technology flourished in some hunter-gatherer groups and not others. Direct evidence of pottery use can be gained through molecular and stable isotope analyses of lipid residues and other organic matter preserved in association with ceramics^{10,11}. Although such investigations have been widely conducted on vessels associated with early agricultural or pastoral communities^{12,13}, the use of earlier hunter-gatherer pottery has been largely ignored or considered only for late forager groups^{14,15}. The main challenge limiting systematic investigations of early hunter-gatherer pottery use is the extent to which molecular evidence is preserved after such long-term exposure to the burial environment. The earliest known pottery use in Eurasia gradually developed, probably independently, in Southern China, the Amur River basin (Russian Far East) and Japan at the end of the Late Pleistocene^{1–5}, necessitating the recovery of organic residues at least 3,000 yr older than has so far been reported¹². The rarity of the very earliest pottery vessels also limits widespread destructive residue analysis. Here we present an investigation of Late Pleistocene pottery use, focusing on the Japanese ceramic sequence known as the Incipient Jōmon, using small sub-samples (~20 mg) of charred deposits adhering to the inside and outside of potsherds. We investigated the carbon and nitrogen stable isotope characteristics of the charred deposits using bulk isotope ratio mass spectrometry (IRMS). Where the sample was large enough, lipids were extracted and analysed by gas chromatography mass spectrometry (GCMS) and GC combustion IRMS (GC-C-IRMS) to determine their structural and carbon isotope characteristics^{10,11}.

Incipient Jōmon pottery is thought to have been invented in Japan ~16,000 cal BP², and became established across the major islands of the Japanese archipelago during the last 4,000 yr of the Late Pleistocene¹⁶. Much is known about the glacial foragers who developed this pottery. Their sites have been well dated¹⁷ and the pottery sequence has been intensively studied⁷: three major phases of typological development have been identified, each broadly mapping onto different climate phases of the final Pleistocene (Fig. 1). Japanese Late Pleistocene pottery containers were produced in only very small quantities (Fig. 1b) and their estimated volumes were also relatively low. Therefore, pottery may have initially been reserved for specialized subsistence activities or ritual use, rather than being a widespread technology integral to everyday culinary practices. Nevertheless, this long-lived nascent phase, which in some regions endured severe climatic fluctuations, led to an extremely rapid expansion in pottery production in the early Holocene, possibly in response to a warming climate, demographic expansion and intensification of resource exploitation.

Charred surface deposits were sampled from 101 Incipient Jōmon vessels from 13 sites, dating from 15,300–11,200 cal BP (Supplementary Table 1) and encompassing a wide range of different environments across the archipelago (Fig. 1a). Potsherds from different phases (II and III; Fig. 1b) broadly corresponding to the Bølling–Allerød interstadial and the Younger Dryas stadial were included. Variations in vessel shapes are known to exist in this period^{7,16} but could not be discerned

¹BioArCh, Department of Archaeology, University of York, Heslington, York YO10 5DD, UK. ²Niigata Prefectural Museum of History, Sekihara 1, Nagaoka, Niigata 940-2035, Japan. ³Archaeological Sciences, University of Bradford, Richmond Road, Bradford BD7 1DP, UK. ⁴Division of Chemistry and Environmental Sciences, School of Science and Environmental Science, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK. ⁵School of Environmental Sciences, Nicholson Building, 4 Brownlow Street, The University of Liverpool, Liverpool L69 3GP, UK. ⁶Research Institute for Humanity and Nature, 457-4 Kamigamo-motoyama, Kita-ku, Kyoto 603-8047, Japan. ⁷Wakasa History and Folklore Museum, Onyu 2-104, Obama, Fukui 917-0241, Japan. ⁸Department of Archaeology, University of Aberdeen, Aberdeen AB24 3UF, UK. ⁹Centre for the Study of Cultural Evolution, Stockholm University, SE-106 91 Stockholm, Sweden. ¹⁰The Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, SE-106 91 Stockholm, Sweden. ¹¹Arctic Centre, University of Groningen, PO Box 716, 9700 AS Groningen, Netherlands.

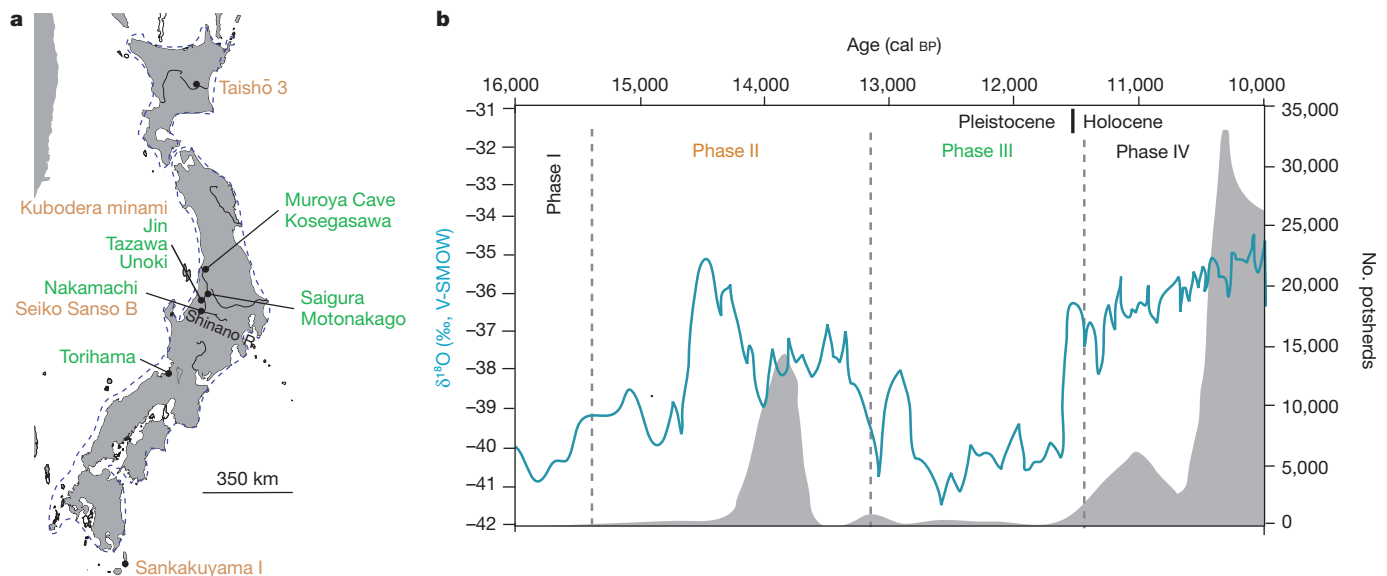


Figure 1 | Locations of sites sampled and temporal changes in early pottery production against Greenland ice-core oxygen isotope record. **a**, Potsherds were selected from assemblages at different sites, dating to chrono-typological phase II, that is, linear-relief wares (orange), and phase III, that is, punctate-marked, nail-impressed and cord-marked wares (green). At Torihama, pottery from both phases was sampled. The coastline corresponding to the Younger

Dryas stadial (~ 65 m below present-day sea level²²) is shown. **b**, The Greenland Ice Sheet Project 2 ice-core $\delta^{18}\text{O}$ values (blue) generally reflect temperature, with lower $\delta^{18}\text{O}$ values corresponding to lower temperatures. Potsherd quantities (grey) are as previously reported^{7,16}, calculated from 52 site reports and chronologically assigned by their typology and by 118 associated ^{14}C dates.

because of the fragmentary nature of the assemblages. Bulk carbon ($\delta^{13}\text{C}$) and nitrogen ($\delta^{15}\text{N}$) isotope ratio values obtained from each charred deposit were variable across the whole sample and between individual sites (Fig. 2a). Despite this variability, more than 75% of the samples had $\delta^{15}\text{N}$ values greater than 9‰ (median, 10.8‰) consistent with the processing of high-trophic-level aquatic (that is, marine or freshwater) products, as has been observed experimentally (Fig. 2a) and on pottery from different archaeological sites associated with the exploitation of fish and marine mammals^{14,18,19}. To confirm this interpretation, lipids were extracted from 57 charred interior deposits and four exterior deposits from Incipient Jōmon vessels and analysed by

GCMS (Table 1). Seventeen samples from the Torihama site, Western Honshu, and one from the Taishō 3 site, Hokkaido, produced distinctive lipid profiles typically consisting of medium- and long-chain saturated (C_{14} – C_{24}) and mono-unsaturated ($\text{C}_{16:1}$ – $\text{C}_{22:1}$) fatty acids, isoprenoid fatty acids and long-chain (C_{18} – C_{22}) ω -(*o*-alkylphenyl) fatty acids (Fig. 3 and Supplementary Table 2). These profiles are typical of degraded aquatic oils, meeting the established criteria for their identification in archaeology²⁰ and providing compelling evidence that freshwater and/or marine foods were processed in these vessels. Such lipid profiles were not observed in the four exterior surface deposits and are highly unlikely to derive from the depositional

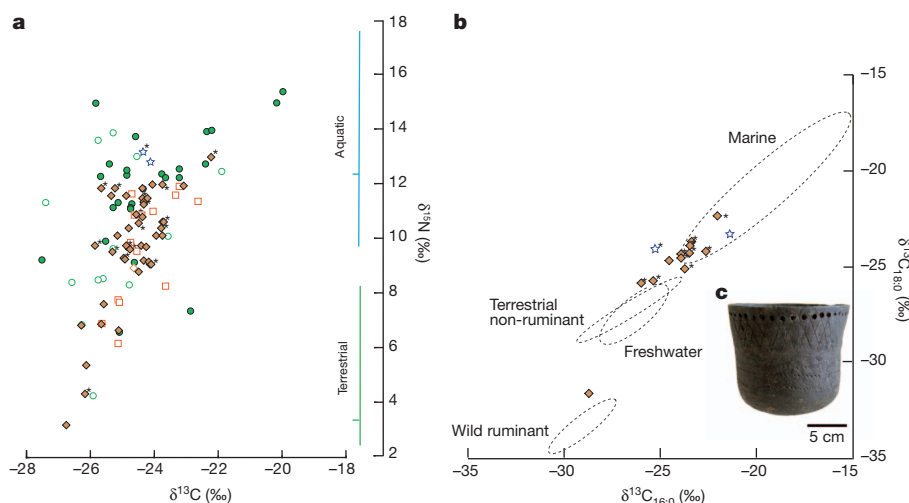


Figure 2 | Bulk carbon and nitrogen stable isotope ratios and fatty acid stable carbon isotope ratios from Incipient Jōmon pottery. **a**, Bulk stable isotope values are discriminated according to site location and period. Torihama, orange diamonds; Shinano river sites, green circles; Sankakuyama I, red squares; Taishō 3, blue stars. Filled symbols indicate phase III and open symbols indicate phase II. The blue and green vertical bars show 95% confidence intervals and the median stable nitrogen isotope value from experimentally charred aquatic (blue, $n = 9$) and terrestrial (green, $n = 8$)

organisms (Supplementary Table 3). **b**, Plot of the $\delta^{13}\text{C}$ values of $\text{C}_{16:0}$ and $\text{C}_{18:0}$ fatty acids from charred deposits on pottery from Torihama and Taishō 3. These are compared with reference fats from wild terrestrial non-ruminant ($n = 5$), wild ruminant ($n = 10$), freshwater ($n = 9$) and marine ($n = 17$) organisms (Supplementary Table 4). Data are plotted with 68% confidence intervals (using STATISTICA v.7). Samples that produced aquatic biomarkers are marked with asterisks. **c**, Photograph of a reconstructed Incipient Jōmon vessel from Torihama.

Table 1 | Summary of GCMS results of charred deposits on Incipient Jōmon pottery

Site	Region	Phase	Associated dates* ($\times 10^3$ cal BP)	No. samples analysed	No. samples yielding lipid†	No. samples yielding aquatic biomarkers
Kubodera minami	Central Honshu	II	15.3–14.2	4	0	0
Taishō 3	Hokkaido	II	15.0–14.0	2	2	1
Seiko Sanso B	Central Honshu	II	14.1–13.9	1	1	0
Sankakuyama I	Kyushu	II	13.6–13.4	13	0	0
Torihamma	Western Honshu	III	12.1–11.8	35	32	17
Muroya Cave	Central Honshu	III	NA	1	0	0
Tazawa	Central Honshu	III	NA	1	0	0

Further details are available in Supplementary Table 2.

*Range corresponds to ± 2 s.d.

†Greater than $0.1 \mu\text{g mg}^{-1}$ of food crust.

NA, not available.

environment because the ω -(*o*-alkylphenyl) alkanolic acids are formed only at high temperatures ($>270^\circ\text{C}$) from the protracted heating of polyunsaturated fatty acids²⁰. The presence of one of these biomarkers in two exterior deposits (Supplementary Table 2) is most probably attributable to overspill of the vessel's contents.

Further information regarding the source of the organic residues was obtained by measuring the stable carbon isotope ratio of two medium-chain-length saturated fatty acids ($\text{C}_{16:0}$ and $\text{C}_{18:0}$) preserved in sufficient quantities in the charred surface deposits adhering to 12 of the Torihamma vessels and two of those from Taishō 3. Eleven samples, including nine with characteristic aquatic biomarkers, yielded either $\text{C}_{16:0}$ or $\text{C}_{18:0}$ fatty acids noticeably enriched in ^{13}C (that is, $\delta^{13}\text{C} \geq -25\text{‰}$). These $\delta^{13}\text{C}$ values are more consistent with authentic marine products than with freshwater or terrestrial animals (Fig. 2b). Two samples from Torihamma, both with aquatic biomarkers, yield $\delta^{13}\text{C}$ values closer to those of lipid extracts from modern pottery vessels used to prepare freshwater as well as non-ruminant animals from Japan (Fig. 2b), but their use cannot be resolved more specifically. In contrast, one Torihamma sample (T66), which was depleted in ^{15}N ($\delta^{15}\text{N} = 6.8\text{‰}$), contained fatty acids with an isotope composition comparable to wild ruminant fats²¹ (Fig. 2b). Aquatic biomarkers were absent from this sample and the lipid profile was instead dominated by saturated fatty acids (C_{14} – C_{18}) that are more typical of degraded terrestrial animal fat, indicating a very different use history of this particular pottery vessel.

Despite some variation in pottery use at Torihamma, the majority of the samples, and all but one with aquatic biomarkers, had relatively high bulk $\delta^{13}\text{C}$ ($> -26\text{‰}$) and $\delta^{15}\text{N}$ ($> 9\text{‰}$) values, the latter consistent with charred deposits formed experimentally from authentic marine organisms (Supplementary Table 3) and both comparable to those associated with Northern European hunter-gatherer pottery found at coastal sites¹⁴. Both the bulk and the compound-specific isotope results from the earlier site of Taishō 3 ($\sim 15,000$ – $14,000$ cal BP) are also consistent with the processing of marine products in pottery vessels (Fig. 2). Therefore, we suggest that the preparation of aquatic resources, particularly marine foodstuffs, was the dominant mode of pottery use at these sites even though neither site is situated directly on the coast (Fig. 1a). At Torihamma, the nearby exposed littoral shelf conceivably provided easy access to a range of marine mammals and fish during the Late Pleistocene, even considering the slightly lower sea level during that period²² (Fig. 1a). An alternative explanation is that migratory fish, such as salmon, were caught much closer to each site as they travelled upriver from marine feeding grounds.

More generally, considering the 13 Incipient Jōmon sites, the distribution of $\delta^{15}\text{N}$ values in charred deposits associated with pots that date from the warmer Bolling–Allerød interstadial, when there was greater deciduous forest cover²³ and larger numbers of large mammals, is not significantly different (Mann–Whitney U test) from that of the Younger Dryas stadial for sites on Honshu ($U = 392$, $P = 0.21$, $n_{\text{Bolling–Allerød}} = 14$, $n_{\text{Younger Dryas}} = 71$) or overall ($U = 893$, $P = 0.20$, $n_{\text{Bolling–Allerød}} = 30$, $n_{\text{Younger Dryas}} = 71$). From both the bulk stable isotope data and the more specific product identifications based on available lipid data, we suggest that aquatic products were the most frequently processed products in Incipient Jōmon pottery, through the

fluctuating climate and across a range of environments, from Hokkaido in the north to Kyushu in the south. Whether ceramic vessels were integral to the processing of aquatic resources or, as is suggested by their rarity, were used only occasionally, perhaps ceremonially or as a prestige technology²⁴, remains debatable. The appearance of pottery at Late Pleistocene sites located on narrowing coastal plains and along rivers where there was an abundance of terrestrial game, aquatic resources and fresh water²⁵ might have led to decreased mobility permitting a greater investment in pottery production for either purpose. The long ‘experimental’ phase of pottery use in the late glacial contrasts

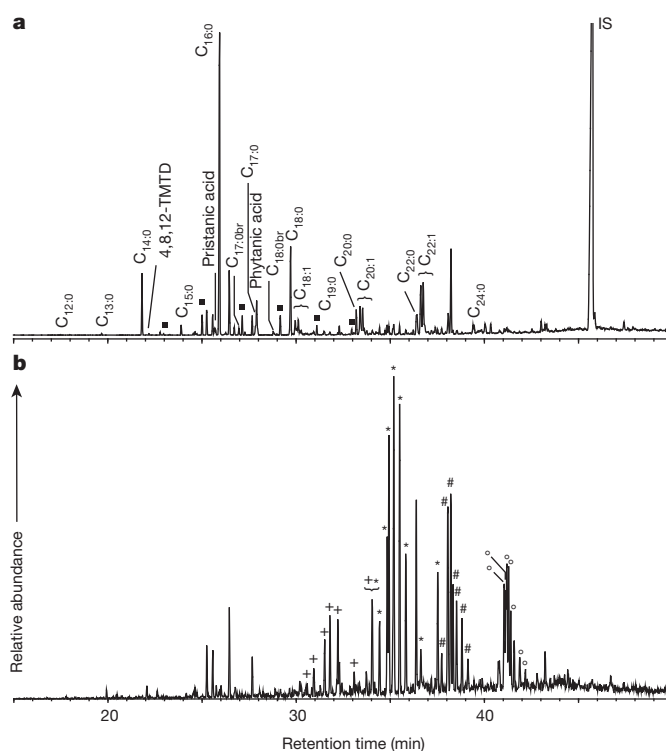


Figure 3 | Partial gas chromatograms of a lipid extract from a charred deposit adhering to a Torihamma potsherd ($\sim 12,000$ cal BP). **a**, The total ion chromatogram is characteristic of degraded aquatic oil and is dominated by medium- and long-chain saturated and mono-unsaturated fatty acids and isoprenoid fatty acids, pristanic, phytanic and 4,8,12-TMTD (4,8,12-trimethyltridecanoic acid). α,ω -dicarboxylic acids (■) with carbon chain ranges of C_8 – C_{13} were also observed and are the likely oxidation products from the complex mixture of long-chain precursor unsaturated fatty acids^{29,30}, derived from freshwater or marine oils, or both. C_{n-x} fatty acids with carbon length n and number of unsaturations x ; br, branched-chain acids; IS, internal standard (n -hexatriacontane). **b**, The summed m/z 91, m/z 105 (mass-to-charge ratio) ion chromatogram shows the presence of ω -(*o*-alkylphenyl) alkanolic acids with 16 (+), 18 (*), 20 (#) or 22 (O) carbon atoms. The presence of the third and fourth components, thermally produced from C_{20} and C_{22} unsaturated fatty acids, and the distribution of their isomers confirm that this residue (sample T124) is derived from an aquatic organism³⁰.

sharply with its sudden ubiquity in the Early Holocene of Japan (Fig. 1b) and other regions of East Asia⁷, when use intensified and became integrated into everyday practice, triggering further long-range dispersals of the new technology into other regions⁶. Although it is tempting to link this change in pottery abundance directly to resource intensification, increased sedentism and the warmer climate, more complex demographic and social factors could also be posited²⁶. Our study, showing that organic residues can be reliably obtained from Late Pleistocene pottery, will make it possible to clarify, through further widespread analysis, the drivers of the continued development of pottery in the Early Holocene of East Asia and its emergence at other innovation centres.

METHODS SUMMARY

Subsamples of charred surface deposits (10–30 mg) were removed with a sterile scalpel from 101 potsherds from 13 different sites (Supplementary Table 3 and Fig. 1a). The chronological phase of each sample was established by its typology, stratigraphic position and ¹⁴C dating of associated materials (Supplementary Table 1). In most cases these agree but, where available, radiocarbon dates of materials directly associated with the pottery analysed were favoured to assign a chronological phase using the published date ranges⁷. Lipids were extracted by alkali saponification with NaOH in methanol²⁷ (5% (w/v), 2 h, 70 °C) or extracted and methylated in one-step with acidified methanol²⁸, or both. For the second process, methanol (1 ml) was added to homogenized charred deposits (10–30 mg) and the mixture was sonicated for 15 min and then acidified with concentrated sulphuric acid (200 µl). The acidified suspension was heated in sealed tubes for 4 h at 70 °C and then cooled, and lipids were extracted with *n*-hexane (3 × 2 ml) and directly analysed by GCMS and GC-C-IRMS using standard protocols^{18,20,21}. Saponified extracts were cooled, neutral lipids were removed (*n*-hexane, 3 × 2 ml), the extracts were acidified with HCl, and the acid fraction was extracted (*n*-hexane, 3 × 2 ml) and methylated to produce fatty acid methyl esters (FAMES) using BF₃-methanol complex (14% (w/v), 200 µl, 1 h, 70 °C). For GC-C-IRMS, instrument precision on repeated measurements was ±0.3‰ (s.e.m.) and the accuracy determined from FAME and *n*-alkane isotope standards was ±0.5‰ (s.e.m.). Modern samples and experimental pots were extracted using exactly the same procedure or as previously reported^{21,27}, or both. In addition, crushed surface residues (~1 mg) were analysed by elemental analysis IRMS as previously reported¹⁸. Samples yielding less than 1% N were discarded and instrument precision on repeated measurements was ±0.2‰ (s.e.m.). $\delta^{13}\text{C}$, $\delta^{15}\text{N}$ or $\delta^{18}\text{O}$ = $[(R_{\text{sample}}/R_{\text{standard}}) - 1] \times 1,000$, where $R = {}^{13}\text{C}/{}^{12}\text{C}$, ${}^{15}\text{N}/{}^{14}\text{N}$ or ${}^{18}\text{O}/{}^{16}\text{O}$. The standard for $\delta^{13}\text{C}$ is Vienna Pee Dee Belemnite (V-PDB), the standard for $\delta^{15}\text{N}$ is air N₂ and the standard for $\delta^{18}\text{O}$ is Vienna Standard Mean Ocean Water (V-SMOW).

Received 31 January; accepted 20 March 2013.

Published online 10 April 2013.

- Wu, X. *et al.* Early pottery at 20,000 years ago in Xianrendong Cave, China. *Science* **336**, 1696–1700 (2012).
- Nakamura, T., Taniguchi, Y., Tsuji, S. & Oda, H. Radiocarbon dating of charred residues on the earliest pottery in Japan. *Radiocarbon* **43**, 1129–1138 (2001).
- Bailey, G. N. & Flemming, N. C. Archaeology of the continental shelf: marine resources, submerged landscapes and underwater archaeology. *Quat. Sci. Rev.* **27**, 2153–2165 (2008).
- Jordan, P. & Zvelebil, M. *Ceramics before Farming: The Dispersal of Pottery Among Prehistoric Eurasian Hunter-Gatherers* (Left Coast, 2009).
- Keally, C. T., Taniguchi, Y., Kuzmin, Y. V. & Shewkomud, I. Y. Chronology of the beginning of pottery manufacture in East Asia. *Radiocarbon* **46**, 345–351 (2004).
- Jordan, P. & Zvelebil, M. in *Ceramics Before Farming: The Dispersal of Pottery Among Prehistoric Eurasian Hunter-Gatherers* (eds Jordan, P. & Zvelebil, M.) 33–90 (Left Coast, 2009).
- Keally, C. T., Taniguchi, Y. & Kuzmin, Y. V. Understanding the beginnings of pottery technology in Japan and neighboring East Asia. *Rev. Archaeol.* **24**, 3–14 (2003).

- Rice, P. M. On the origins of pottery (ceramic technology among complex hunter-gatherers). *J. Archaeol. Method Theory* **6**, 1–54 (1999).
- Ikawa-Smith, F. On ceramic technology in East Asia. *Curr. Anthropol.* **17**, 513–515 (1976).
- Evershed, R. P. Organic residue analysis in archaeology: the archaeological biomarker revolution. *Archaeometry* **50**, 895–924 (2008).
- Regert, M. Analytical strategies for discriminating archaeological fatty substances from animal origin. *Mass Spectrom. Rev.* **30**, 177–220 (2011).
- Evershed, R. P. *et al.* Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* **455**, 528–531 (2008).
- Craig, O. E. *et al.* Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity* **79**, 882–894 (2005).
- Craig, O. E. *et al.* Ancient lipids reveal continuity in culinary practices across the transition to agriculture in Northern Europe. *Proc. Natl Acad. Sci. USA* **108**, 17910–17915 (2011).
- Isaksson, S. Vessels of change: a long-term perspective on prehistoric pottery-use in southern and eastern middle Sweden based on lipid residue analyses. *Curr. Swedish Archaeol.* **17**, 131–149 (2009).
- Taniguchi, Y. The beginning of the early Jōmon. *Ibun* **20**, 2–36 (2002).
- Ono, A., Sato, H., Tsutsumi, T. & Kudo, Y. Radiocarbon dates and archaeology of the Late Pleistocene in the Japanese islands. *Radiocarbon* **44**, 477–494 (2002).
- Craig, O. E. *et al.* Molecular and isotopic demonstration of the processing of aquatic products in northern European prehistoric pottery. *Archaeometry* **49**, 135–152 (2007).
- Morton, J. D. & Schwarcz, H. P. Palaeodietary implications from stable isotopic analysis of residues on prehistoric Ontario ceramics. *J. Archaeol. Sci.* **31**, 503–517 (2004).
- Hansel, F. A., Copley, M. S., Madureira, L. A. S. & Evershed, R. P. Thermally produced ω -(*o*-alkylphenyl) alkanolic acids provide evidence for the processing of marine products in archaeological pottery vessels. *Tetrahedr. Lett.* **45**, 2999–3002 (2004).
- Craig, O. E. *et al.* Distinguishing wild ruminant lipids by gas chromatography/combustion/isotope ratio mass spectrometry. *Rapid Commun. Mass Spectrom.* **26**, 2359–2364 (2012).
- Tanabe, S., Nakanishi, T. & Yasui, S. Relative sea-level change in and around the Younger Dryas inferred from late Quaternary incised-valley fills along the Japan Sea. *Quat. Sci. Rev.* **29**, 3956–3971 (2010).
- Kudo, Y. & Kumon, F. Paleolithic cultures of MIS 3 to MIS 1 in relation to climate changes in the central Japanese islands. *Quat. Int.* **248**, 22–31 (2012).
- Hayden, B. in *The Emergence of Pottery* (eds Barnett, W. K. & Hoopes, J. W.) 257–266 (Smithsonian Institution, 1995).
- Faure, H., Walter, R. C. & Grant, D. R. The coastal oasis: ice age springs on emerged continental shelves. *Global Planet. Change* **33**, 47–56 (2002).
- Kaner, S. in *Ceramics Before Farming: The Dispersal of Pottery Among Prehistoric Eurasian Hunter-Gatherers* (eds Jordan, P. & Zvelebil, M.) 93–119 (Left Coast, 2009).
- Craig, O. E., Love, G. D., Isaksson, S., Taylor, G. & Snape, C. E. Stable carbon isotopic characterisation of free and bound lipid constituents of archaeological ceramic vessels released by solvent extraction, alkaline hydrolysis and catalytic hydropyrolysis. *J. Anal. Appl. Pyrolysis* **71**, 613–634 (2004).
- Hamilton, R. J., Hamilton, S. & Harwood, J. *Lipid Analysis: A Practical Approach* 53–54 (IRL, 1992).
- Regert, M., Bland, H. A., Dudd, S. N., van Bergen, P. F. & Evershed, R. P. Free and bound fatty acid oxidation products in archaeological ceramic vessels. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 2027–2032 (1998).
- Evershed, R. P., Copley, M. S., Dickson, L. & Hansel, F. A. Experimental evidence for the processing of marine animal products and other commodities containing polyunsaturated fatty acids in pottery vessels. *Archaeometry* **50**, 101–113 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Leverhulme trust (F/00 152/AM) and the Japanese Society for the Promotion of Science (PE 11560) for their support. We are grateful to K. Adachi, K. Higashi, Y. Kasai, H. Kato, K. Nagahama, H. Oguma, T. Tsuchiya, T. Watanabe and T. Yamahara for providing access to samples.

Author Contributions O.E.C., H.S., Y.N., S.I. and P.J. planned the project. O.E.C. wrote the paper with assistance from all other authors. P.J., H.S., Y.N. and K.G. carried out sampling with assistance of M.A. and J.U., who provided contextual data. O.E.C., H.S., A.L., K.T., D.A. and A.T. carried out the lipid analysis. C.P.H. and L.C. carried out the bulk stable isotope analyses. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.E.C. (oliver.craig@york.ac.uk).

Slower recovery in space before collapse of connected populations

Lei Dai¹, Kirill S. Korolev¹ & Jeff Gore¹

Slower recovery from perturbations near a tipping point and its indirect signatures in fluctuation patterns have been suggested to foreshadow catastrophes in a wide variety of systems^{1,2}. Recent studies of populations in the field and in the laboratory have used time-series data to confirm some of the theoretically predicted early warning indicators, such as an increase in recovery time or in the size and timescale of fluctuations^{3–6}. However, the predictive power of temporal warning signals is limited by the demand for long-term observations. Large-scale spatial data are more accessible, but the performance of warning signals in spatially extended systems^{7–10} needs to be examined empirically^{3,11–13}. Here we use spatially extended yeast populations, an experimental system with a fold bifurcation (tipping point)⁶, to evaluate early warning signals based on spatio-temporal fluctuations and to identify a novel spatial warning indicator. We found that two leading indicators based on fluctuations increased before collapse of connected populations; however, the magnitudes of the increases were smaller than those observed in isolated populations, possibly because local variation is reduced by dispersal. Furthermore, we propose a generic indicator based on deterministic spatial patterns, which we call ‘recovery length’. As the spatial counterpart of recovery time¹⁴, recovery length is the distance necessary for connected populations to recover from spatial perturbations. In our experiments, recovery length increased substantially before population collapse, suggesting that the spatial scale of recovery can provide a superior warning signal before tipping points in spatially extended systems.

Positive feedback is widespread in nature, ranging from cellular circuits to population growth to the melting of ice sheets. There is growing evidence that positive feedback leads to alternative stable states and tipping points in various ecological systems^{15–18}. Closer to a tipping point an ecosystem becomes less resilient and more likely to shift to an alternative state¹⁹ such as the collapse of fish stocks, eutrophication of lakes, or loss of vegetation²⁰. Predicting these undesirable transitions may sound like an impossible task because of the inherent complexity underlying these systems. However, recent advances incorporating ideas from nonlinear dynamical systems theory suggest that there may be signatures of “critical slowing down” in the vicinity of tipping points^{1,2}. At the brink of these sudden transitions, the recovery of a system after perturbations should slow down¹⁴, also leading to changes in the pattern of fluctuations²¹. Thus, a set of indicators related to critical slowing down may provide advance warning of an impending transition. Empirical tests in the field⁴ and in the laboratory^{3,5,6} have revealed some of the early warning signals based on fluctuations in time series, such as temporal variation and autocorrelation.

However, our understanding of early warning signals in spatially extended systems is still limited^{1,2}. The studies in time series typically ignore spatial interactions; in reality spatial coupling between habitat patches (for example, dispersal of populations or exchange of biomass) is common and may affect the performance of some warning signals²². Moreover, temporal warning signals rely on data from long-term observations, which are scarce and difficult to obtain. Large-scale spatial data, such as satellite-derived data sets¹⁷, could be more readily available.

Spatial data not only provide a greater quantity of information, but also allow us to study features of the system that are not available through time series. Statistical indicators based on spatial fluctuations have been proposed^{7–10} but empirical studies are limited^{3,11,12}; tests of these indicators in replicated experiments, which avoid the bias introduced by selective sampling²³, are lacking. In addition, previous studies of vegetation systems discovered emerging spatial patterns preceding transitions^{24,25}. However, the vegetation patterns are often specific to the system studied, whereas identifying generic spatial warning signals would add a powerful tool to the analysis of ecosystem stability. Here we address these questions using an experimental system of spatially extended yeast populations with alternative stable states and a tipping point leading to population collapse.

We grew laboratory populations of the budding yeast *Saccharomyces cerevisiae* in sucrose and performed daily dilution into fresh media. During the daily dilution, a fraction (for example, 1 in 500 for a dilution factor of 500) of the cells were transferred to fresh media. This is a well characterized system with an experimentally mapped fold bifurcation (tipping point)⁶. Yeast cells grow cooperatively in sucrose by sharing the hydrolysis products²⁶, creating a positive feedback between cells that leads to bistability and a tipping point (Supplementary Fig. 1). By increasing the dilution factor (equivalent to an increase in the mortality rate), we could drive isolated yeast populations to collapse on crossing the tipping point (Fig. 1a).

We then connected local yeast populations spatially through controlled dispersal between nearest neighbours on a one-dimensional array (Fig. 1b). Spatial coupling between local populations was introduced by adding a dispersal step during the daily dilution. In the dispersal step, 25% (corresponding to a dispersal rate $D = 2 \times 25\% = 0.5$) of a local population was transferred to each of its nearest neighbours; the rest of the population remained in the patch. For each dilution factor, there were four replicate arrays each consisting of ten patches. A group of isolated populations ($D = 0$) was grown in a similar experimental setting except that there was no mixing between neighbours (Methods). The isolated populations served as a control group and allowed us to investigate the effects of spatial coupling on warning signals. From dilution factor 500 to 1,600, both groups of connected and isolated populations survived and reached equilibrium densities in a week; at dilution factor 1,700, most of the populations collapsed within the timescale of our experiment (insets to Fig. 1a).

After the populations stabilized, we tracked the fluctuations of population density around equilibrium for at least five days to calculate statistical indicators (Methods). Consistent with critical slowing down, we observed a clear increase in the coefficient of variation (CV) of connected populations towards the tipping point (Fig. 2a); however, the magnitude of increase in CV was smaller than in the isolated populations. We then tested lag-1 autocorrelation, a leading indicator for the temporal correlation of fluctuations. As expected, we found that the temporal correlation of connected populations increased gradually to around 0.6 in the vicinity of the tipping point (Fig. 2b). Similar to the observation in CV, the signal in temporal correlation was weaker than in the isolated populations. Although fluctuations of population density

¹Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

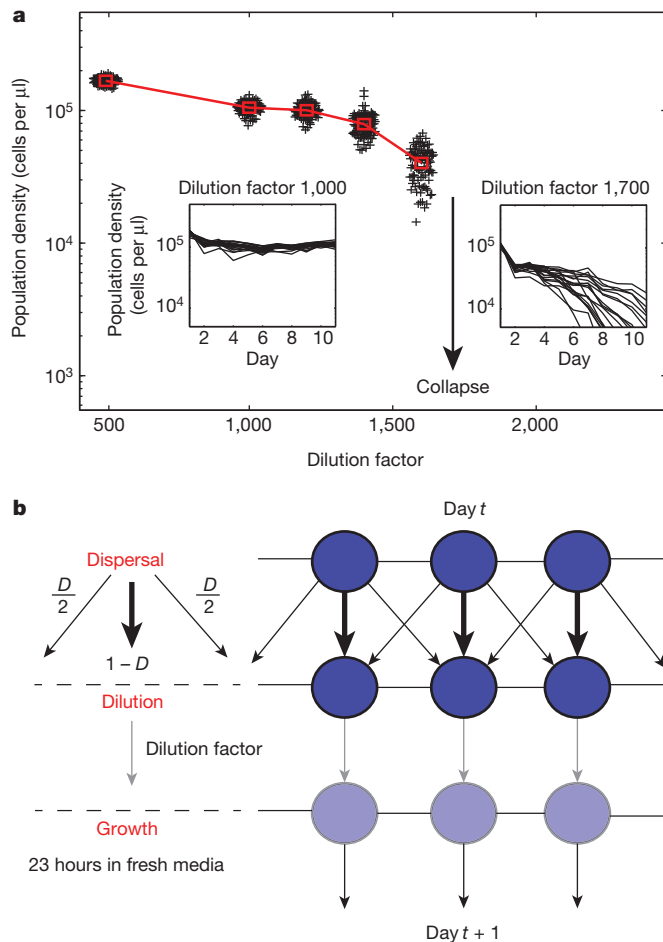


Figure 1 | Yeast populations with a tipping point: an experimental system to study the collapse of connected populations. **a**, Isolated yeast populations collapse after crossing a tipping point. The distribution of population density around equilibrium is shown in spread points; the red square denotes the mean. Insets are traces of replicate populations at dilution factor 1,000 (stable) and 1,700 (collapsed). **b**, Yeast populations are spatially connected by controlled daily dispersal. Each circle corresponds to a habitat patch where a local population grows. A fraction of the local population is transferred to each of its two nearest neighbours, and the rest to itself.

in general became larger and more correlated before population collapse, we found that these two warning signals (CV and lag-1 autocorrelation) seemed to be suppressed in the presence of dispersal, especially at higher dilution factors.

One explanation for the observed suppression of the two leading indicators in connected populations is that flows between neighbours smooth out the fluctuations across different patches and effectively reduce the autocorrelation in any local population. Reduced size or timescale of fluctuations due to dispersal among populations was predicted in previous theoretical studies of spatially explicit ecological models^{8,10,22,27}. We note that the smaller warning signals of connected populations in our experiment may be partly due to a minor shift in the tipping point (Supplementary Fig. 2). The averaging effect of dispersal was also found in an independent group of populations subject to '100% dispersal treatment', in which we mixed ten populations completely each day during the dispersal step. In this extreme scenario, the populations showed almost no increase in variation before the tipping point (Supplementary Fig. 3). Moreover, we demonstrated the suppression of CV and lag-1 autocorrelation by dispersal in analytical derivations based on a spatially explicit first-order autoregressive model (Supplementary Note 1) and in stochastic simulations using a phenomenological model of yeast growth⁶ (Supplementary Fig. 4).

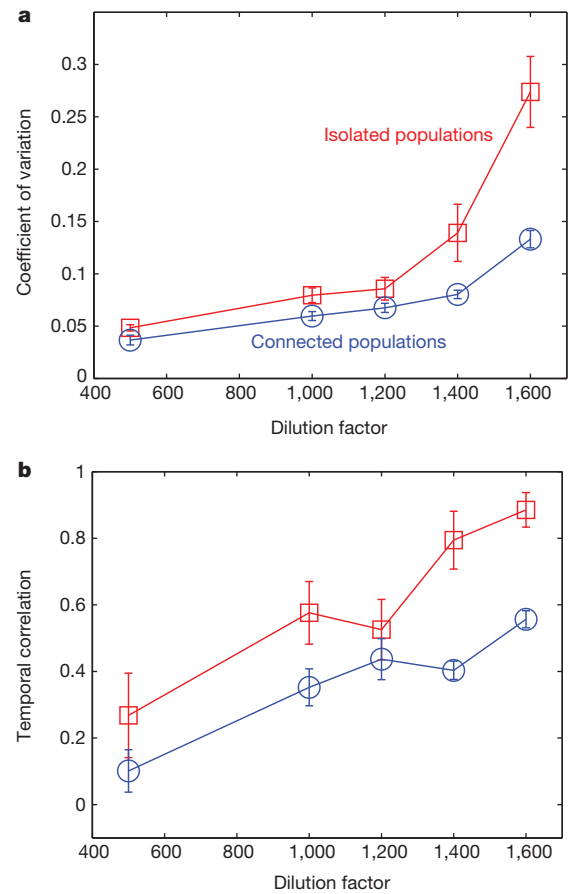


Figure 2 | Early warning signals based on fluctuations show suppressed increase in connected populations. **a**, Coefficient of variation (CV). **b**, Temporal correlation (lag-1 autocorrelation). The coefficient of variation and temporal correlation of both isolated populations (red squares) and connected populations (blue circles) increased before the tipping point. The signals were suppressed in the connected populations, possibly owing to the averaging effect of dispersal. Error bars are standard errors given by bootstrap for isolated populations and standard errors of the mean ($n = 4$) for connected populations.

Spatial coupling introduces the possibility of another warning indicator based on spatial fluctuations: spatial correlation. Long-range spatial correlation has been known to occur in the vicinity of some phase transitions²⁸; recent theoretical work in spatially explicit ecological models found that increasing spatial correlation could be a warning signal before transitions to an alternative stable state⁸. We tested the two-point correlation between nearest neighbours in the connected populations but failed to observe any increase near the tipping point (Supplementary Fig. 5). Simulation results with varying sample size showed that no statistically significant increase in spatial correlation should be discerned with the limited samples in our experiment. Thus, our results suggest that to observe the increase in spatial correlation may require more data than for other indicators.

Facing the potential difficulty of observing a strong warning signal based on fluctuations in spatially connected populations, we set out to look for possible new indicators. The existing warning signals can be classified into different categories, based on the nature of the perturbations and measurements (see Fig. 3). Measuring the recovery time after a pulse perturbation (Fig. 3a) can provide a robust indicator of the distance to a tipping point^{5,14}. In large complex systems, it is often impractical to perform such temporal perturbations repeatedly and measure recovery time. However, owing to stochastic perturbations such as demographic noise, population density constantly fluctuates around the equilibrium. Changes of fluctuation patterns such as an

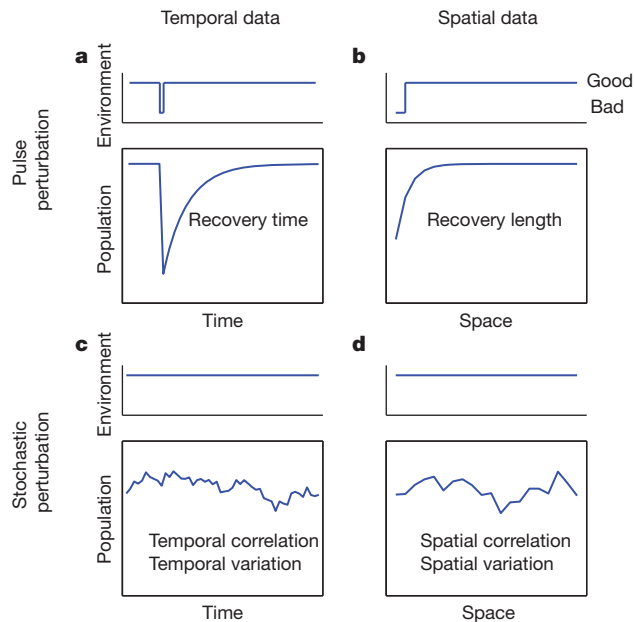


Figure 3 | Early warning signals can be classified into four categories by the nature of perturbations and measurements. **a**, Recovery time; **b**, Recovery length; **c**, Statistical indicators based on temporal fluctuations; **d**, Statistical indicators based on spatial fluctuations. The unexplored category of early warning signals is the spatial counterpart of recovery time: ‘recovery length’. The recovery length characterizes the spatial scale over which population density recovers from a pulse perturbation in space, such as at a boundary with a region of lower quality (**b**). The recovery length increases towards the tipping point (Supplementary Note 2) and provides a novel indicator of critical slowing down in spatial data.

increase in variation and correlation (Fig. 3c, d), measured either in time or in space, are also signatures of critical slowing down and consist of another two categories of leading warning signals^{3,4,6–10,21}. Surprisingly, there is one remaining category that has not been proposed: find or create a ‘pulse perturbation in space’ (Fig. 3b) and measure the spatial counterpart of recovery time¹⁴. Adjacent to a region of poor quality, the neighbouring good patches will not immediately have reached their carrying capacity; instead the carrying capacity will be reached only further from the bad region (Supplementary Fig. 6). Rather than an increase in the timescale to recover, critical slowing down here manifests itself as an increase in the spatial scale to recover (Supplementary Note 2), that is, an increase in ‘recovery length’ as compared to ‘recovery time’.

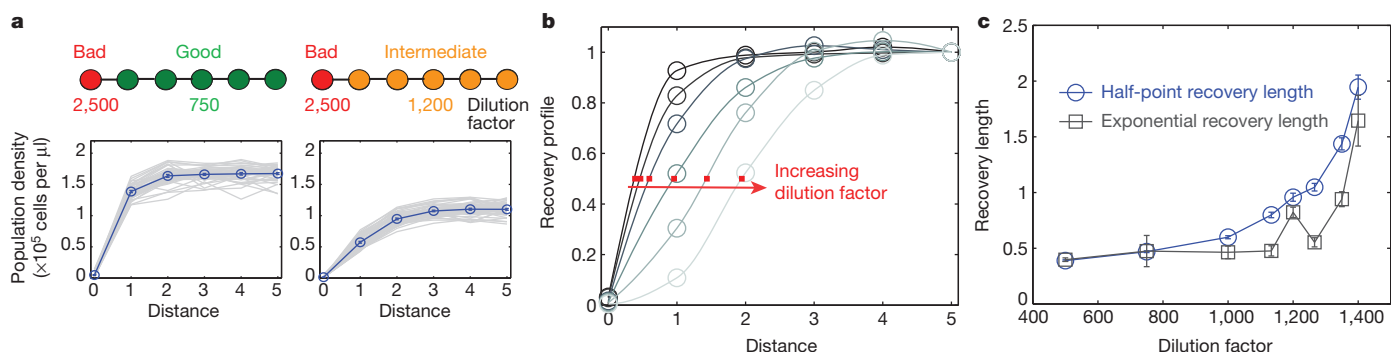


Figure 4 | Recovery length provides a direct measure of critical slowing down in space. **a**, Connected populations in relatively good regions of dilution factor 750 (left) and 1,200 (right) recover gradually in space to the equilibrium density. Blue circles denote the steady-state profile of population density after averaging over replicates (shown in grey). **b**, Recovery profiles at dilution factor 500, 750, 1,000, 1,200, 1,350 and 1,400 show an increasing spatial scale of

recovery. The profile is normalized by the population density of the patch furthest from the bad region. Lines are shape-preserving interpolations; the position of half-recovery is marked by a red square. **c**, Two different measures of recovery length increase substantially with dilution factor. Error bars are standard errors given by bootstrap.

To test our ‘recovery length’ hypothesis, we performed another set of experiments with spatially connected yeast populations (dispersal rate $D = 0.5$), now with two different regions: a relatively good (lower dilution factor) region of five patches and a bad (high dilution factor) region of one patch (Fig. 4a). Given this sharp boundary between two regions of different quality, population density in the good region recovered gradually in space to the equilibrium value. As the condition of the good region deteriorates, we expect an increase in the spatial scale over which the populations recover. Indeed, we observed a clear change in the steady-state recovery profile of populations with increasing dilution factor of the good region (Fig. 4b). In agreement with our hypothesis, the spatial recovery spanned a much longer distance closer to the tipping point.

We quantified this spatial scale using two different indicators (Fig. 4c). The first indicator, the ‘half-point recovery length’, measures the distance between the bad region and the location of half recovery (Methods). The half-point recovery length increased gradually with dilution factor from less than 0.5 to around 2. The second indicator, the ‘exponential recovery length’, is obtained by fitting the recovery profile with an exponential function (Methods). Similar to the first indicator, the exponential recovery length increased more than three-fold as the tipping point was approached. Thus, both measures suggest that the recovery length provides a strong warning signal before population collapse in our system. We also observed an increase in both indicators as we slowly caused the good region to deteriorate by increasing the dilution factor and induced the collapse of connected populations in real time (Supplementary Fig. 7).

Recovery length completes the four categories of early warning signals and can help improve our assessment of spatially extended systems. Our results suggest that stronger spatial coupling (higher dispersal rate) suppresses early warning signals in CV and temporal correlation (Supplementary Fig. 4). In contrast, the magnitude of recovery length increases with the level of spatial coupling (Supplementary Note 2). These two categories of early warning signals are therefore complementary: when one signal is weak the other is strong. Also, although our experiment was conducted on a linear array, the use of recovery length can be readily generalized to two-dimensional systems by mapping the profile perpendicularly to contours of population density. Unlike the specific spatial patterns found in two-dimensional vegetation systems^{24,25}, recovery length may provide a generic measure given that the spatially coupled units by themselves would recover more slowly near the tipping point.

Finally, from a practical perspective, boundaries between regions of different quality are ubiquitous in nature, thus providing many opportunities to measure the recovery length in populations of interest. One specific example of recovery length would be the ‘distance of

edge influence" in landscape ecology²⁹: it quantifies the spatial scale of edge influence on biota in fragmented landscapes. Data of edge influence for forests at different sites suggest a longer recovery length in the Australian tropical forests³⁰, which seems to be borne out by the recent forest collapses in Western Australia. In principle, the recovery length can also be measured when spatially extended populations 'recover' from a region of higher quality (Supplementary Figs 6 and 8), suggesting that boundaries may be introduced by conservation efforts (for example, setting up marine reserves).

Our experiments were performed in the simplest spatial setting possible: homogeneous environments and dispersal rates, a large population size and a safe distance away from the tipping point. In the presence of environmental heterogeneity, measurement of recovery length may fail if the desired sharp boundary between regions of different quality is blurred. However, in such a case we would expect enhanced signals in spatial correlation⁸ and spatial variation before population collapse (Supplementary Fig. 9). Our experiments have also not explored the effects of spatial coupling on the global stability of a meta-population. On the one hand, spatial coupling may reduce fluctuations and the probability that a random shock will trigger a state shift²²; on the other hand, stochastic local extinctions or the introduction of a bad region may drive the connected populations to collapse before the tipping point of a local population (Supplementary Note 3).

Our work illustrates the important role of spatial coupling, such as the dispersal of populations, in our understanding of how to apply the current toolbox of warning indicators to natural populations. More empirical studies are required to confirm the generality and applicability of different indicators, but being able to observe warning signals in connected populations suggests that we may be able to develop quantitative metrics for assessing the fragility of spatially extended complex systems.

METHODS SUMMARY

We grew the budding yeast *Saccharomyces cerevisiae* in 200 µl batch culture on BD Falcon 96-well Microtest plates at 30.7 °C (± 0.2 °C, standard deviation) using synthetic media supplemented with 2% sucrose⁶. Serial dilutions were performed daily with variable dilution factors. Population densities were recorded each day before the serial dilution by measuring optical density at 600 nm. Statistical indicators were calculated after the populations stabilized. The coefficient of variation was calculated as the sample standard deviation divided by the sample mean. The temporal correlation was estimated by the Pearson's correlation coefficient between the population densities at subsequent days. In the experiment to measure recovery length, the half-point recovery length was estimated by performing a shape-preserving interpolation to the steady-state recovery profile of population density and then locating the position of half-recovery. The exponential recovery length was estimated by fitting an exponential function to the recovery profile. For further details, see the online-only Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 4 November 2012; accepted 15 March 2013.

Published online 10 April 2013.

1. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
2. Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338**, 344–348 (2012).
3. Drake, J. M. & Griffen, B. D. Early warning signals of extinction in deteriorating environments. *Nature* **467**, 456–459 (2010).
4. Carpenter, S. R. *et al.* Early warnings of regime shifts: a whole-ecosystem experiment. *Science* **332**, 1079–1082 (2011).
5. Veraart, A. J. *et al.* Recovery rates reflect distance to a tipping point in a living system. *Nature* **481**, 357–359 (2012).

6. Dai, L., Vorselen, D., Korolev, K. S. & Gore, J. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* **336**, 1175–1177 (2012).
7. Guttal, V. & Jayaprakash, C. Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems. *Theor. Ecol.* **2**, 3–12 (2009).
8. Dakos, V., Nes, E. H., Donangelo, R., Fort, H. & Scheffer, M. Spatial correlation as leading indicator of catastrophic shifts. *Theor. Ecol.* **3**, 163–174 (2010).
9. Dakos, V., Kéfi, S., Rietkerk, M., Van Nes, E. H. & Scheffer, M. Slowing down in spatially patterned ecosystems at the brink of collapse. *Am. Nat.* **177**, E153–E166 (2011).
10. Carpenter, S. R. & Brock, W. A. Early warnings of regime shifts in spatial dynamics using the discrete Fourier transform. *Ecosphere* **1**, art10 (2010).
11. Lindegren, M. *et al.* Early detection of ecosystem regime shifts: a multiple method evaluation for management application. *PLoS ONE* **7**, e38410 (2012).
12. Litzow, M. A., Urban, J. D. & Laurel, B. J. Increased spatial variance accompanies reorganization of two continental shelf ecosystems. *Ecol. Appl.* **18**, 1331–1337 (2008).
13. Ouyang, Q. & Swinney, H. L. Transition from a uniform state to hexagonal and striped Turing patterns. *Nature* **352**, 610–612 (1991).
14. van Nes, E. H. & Scheffer, M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am. Nat.* **169**, 738–747 (2007).
15. May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
16. Scheffer, M., Carpenter, S. & Foley, J. A. Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
17. Staver, A. C., Archibald, S. & Levin, S. A. The global extent and determinants of savanna and forest as alternative biome states. *Science* **334**, 230–232 (2011).
18. Isbell, F., Tilman, D., Polasky, S., Binder, S. & Hawthorne, P. Low biodiversity state persists two decades after cessation of nutrient enrichment. *Ecol. Lett.* <http://dx.doi.org/10.1111/ele.12066> (2013).
19. Holling, C. S. Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.* **4**, 1–23 (1973).
20. Scheffer, M. *Critical Transitions in Nature and Society* (Princeton Univ. Press, 2009).
21. Kleinen, T., Held, H. & Petschel-Held, G. The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation. *Ocean Dyn.* **53**, 53–63 (2003).
22. Brock, W. A. & Carpenter, S. R. Interacting regime shifts in ecosystems: implication for early warnings. *Ecol. Monogr.* **80**, 353–367 (2010).
23. Boettiger, C. & Hastings, A. Early warning signals and the prosecutor's fallacy. *Proc. R. Soc. Lond. B* **279**, 4734–4739 (2012).
24. Rietkerk, M., Dekker, S. C., De Ruiter, P. C. & Van de Koppel, J. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* **305**, 1926–1929 (2004).
25. Kéfi, S. *et al.* Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems. *Nature* **449**, 213–217 (2007).
26. Gore, J., Youk, H. & Van Oudenaarden, A. Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**, 253–256 (2009).
27. Fernández, A. & Fort, H. Catastrophic phase transitions and early warnings in a spatial ecological model. *J. Stat. Mech.* P09014 (2009).
28. Sole, R. V., Manrubia, S. C., Luque, B., Delgado, J. & Bascompte, J. Phase transitions and complex systems. *Complexity* **1**, 13–26 (1996).
29. Ries, L., Fletcher, R. J., Battin, J. & Sisk, T. D. Ecological responses to habitat edges: mechanisms, models, and variability explained. *Annu. Rev. Ecol. Syst.* **35**, 491–522 (2004).
30. Harper, K. A. *et al.* Edge influence on forest structure and composition in fragmented landscapes. *Conserv. Biol.* **19**, 768–782 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank D. Vorselen, T. Krieger, D. Seekell, M. Pace and members of the Gore laboratory (A. Sanchez, M. Datta, E. Yurtsev, T. Artemova, K. Axelrod and A. Chen) for comments on the manuscript. T. Krieger performed initial simulations for the connected populations. Y. Zhang and O. Ornek collected preliminary data for the experiment to measure recovery length. This work was supported by a Whitaker Health Sciences Fund Fellowship (to L.D.), a Pappalardo Fellowship (to K.S.K.), an NIH R00 Pathways to Independence Award (NIH R00 GM085279-02), an NIH New Innovator Award (NIH DP2), an NSF CAREER Award, a Sloan Research Fellowship, the Pew Scholars Program and the Allen Investigator Program.

Author Contributions L.D., K.S.K. and J.G. designed the study. L.D. performed the experiments and analysis. K.S.K. and J.G. assisted with the analysis. L.D., K.S.K. and J.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.D. (daillei@mit.edu) or J.G. (gore@mit.edu).

METHODS

Experimental protocols. We grew the budding yeast *Saccharomyces cerevisiae* in 200 μ l batch culture on BD Falcon 96-well Microtest plates at 30.7 °C (± 0.2 °C, standard deviation) using synthetic media (yeast nitrogen bases + nitrogen, Complete Supplement Mixture) supplemented with 2% sucrose⁶. Cultures were maintained in a well-mixed condition by growing on a shaker at 825 r.p.m. Serial dilutions were performed daily (23 h of growth) with variable dilution factors. Population densities were recorded each day before the serial dilution by measuring optical density at 600 nm using a Thermo Scientific Varioskan Flash Multimode Reader. The calibration between optical density and cell density was based on the previous characterization of this system⁶.

In the group of connected populations, for each dilution factor there were four replicate arrays each consisting of ten patches. Populations were connected by controlled dispersal between nearest neighbours (dispersal rate $D = 0.5$, which is defined as the fraction of population going out of a patch). Reflecting boundary conditions were adopted, meaning that a population on the edge would have 75% of its cells remaining in the patch during the dispersal step. In the group of isolated populations, the experiment was performed in a similar spatial setting except that there was no dispersal ($D = 0$); for each dilution factor there were four arrays each consisting of five patches, giving a total of 20 replicate populations isolated from each other. The dilution factors for the data presented in Fig. 2 are 500, 1,000, 1,200, 1,400 and 1,600. In the experiment to measure recovery length, populations were connected by nearest-neighbour dispersal ($D = 0.5$, reflecting boundary conditions). The dilution factor for the bad region (one patch) was 2,500; the dilution factor for the good region (five patches) was varied as the environmental driver. The dilution factors for the data presented in Fig. 4 are 500, 750, 1,000, 1,133, 1,200, 1,266, 1,350 and 1,400.

Calculation of statistical indicators. Statistical indicators for the connected populations were calculated among ten populations in one array on each day and averaged over a span of at least five days, after the populations stabilized. The mean value of four replicate arrays and the standard error of the mean ($n = 4$) are shown in Fig. 2. For the isolated populations, statistical indicators were calculated on each day among 20 populations over five days. We used bootstrap to compute the standard errors of the indicators by resampling 1,000 times the ensemble of replicate populations (for the coefficient of variation and the temporal correlation) or arrays (for the spatial correlation).

The coefficient of variation (CV) was calculated as the sample standard deviation (Supplementary Fig. 3c) divided by the sample mean. Because the local populations in our experiment were grown in a homogeneous environment, in principle they could all be treated as replicates. Assuming the system is ergodic, the CV calculated over an ensemble of replicates can be interpreted either as the spatial CV of many populations at one time point or the temporal CV of a single population over many time points. The temporal correlation, defined as the lag-1 autocorrelation, was estimated by the Pearson's correlation coefficient between the population densities at subsequent days. To correct for negative bias in small samples, we used a modified estimator with an additional term $1/N$ for lag-1 autocorrelation³¹. The sample size $N = 10$ for connected populations and

$N = 20$ for isolated populations. N is a fixed number for different dilution factors, so using the modified estimators would not affect the trend of indicators. The spatial correlation, defined as the two-point correlation between all neighbouring pairs, was estimated by the Moran's coefficient^{8,32}. The expectation of Moran's coefficient is $-1/(N-1)$ in the absence of spatial correlation³³; we used a modified estimator with an additional term $1/(N-1)$ so that the expectation is 0. In this case, the sample size N is the number of patches in an array: $N = 10$ for connected populations and $N = 5$ for isolated populations. For detailed formulae of the statistical indicators, see Supplementary Note 4.

In the analysis we ensured environmental homogeneity by removing the linear gradient of population density observed in connected populations. This small gradient is presumably caused by some heterogeneity in experimental conditions (temperature, dilution errors, and so on) across the plate. Removing gradient-type spatial heterogeneity before statistical analysis is similar to the detrending procedure commonly used in time-series analysis; it prevents spurious signals such as positive spatial correlation (Supplementary Fig. 10).

Recovery length. After the recovery profile stabilized, we tracked the population density profiles of at least six replicates over several days. The half-point recovery length L_{half} was estimated by performing a shape-preserving interpolation (Matlab function PCHIP, piecewise cubic Hermite interpolating polynomial) to the recovery profile and then locating the position of half-recovery, at which $n(x = L_{\text{half}}) = (1/2)n(x = 5)$. The population density of the bad region (dilution factor 2,500) in our experiment was close to 0 (Fig. 4 and Supplementary Fig. 7). In the more general scenario with a sharp boundary between two regions of different quality (Supplementary Figs 6 and 8), the position of half-recovery can be defined as the midpoint between the equilibrium population density of the region of interest and the population density at the boundary.

The exponential recovery length L_{exp} was estimated by fitting an exponential function with three parameters $c_1 \exp(-x/L_{\text{exp}}) + c_2$ to the recovery profile $n(x)$. The data points used for exponential fitting are from positions 1 to 5 (except for dilution factor 500, the data for fitting are from positions 0 to 5). We note that our definition of exponential recovery length is phenomenological, because: (1) the deviation is expected to be exponential only close enough to the equilibrium; (2) at higher dilution factors the profile can deviate from an exponential form (Supplementary Fig. 11). The 'kink' in the fitted exponential recovery length (Fig. 4c) may be due to the limited data points used in fitting or experimental errors. For both the half-point recovery length and the exponential recovery length, we used bootstrap to compute standard errors for the indicators by resampling the ensemble of steady-state profiles 100 times and fitting the average recovery profile.

31. DeCarlo, L. T. & Tryon, W. W. Estimating and testing autocorrelation with small samples: a comparison of the c-statistic to a modified estimator. *Behav. Res. Ther.* **31**, 781–788 (1993).
32. Legendre, P. & Fortin, M. J. Spatial pattern and ecological analysis. *Vegetatio* **80**, 107–138 (1989).
33. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).

Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking

Billy T. Chen¹, Hau-Jie Yau¹, Christina Hatch¹, Ikue Kusumoto-Yoshida¹, Saemi L. Cho², F. Woodward Hopf^{2,3} & Antonello Bonci^{1,3,4}

Loss of control over harmful drug seeking is one of the most intractable aspects of addiction, as human substance abusers continue to pursue drugs despite incurring significant negative consequences¹. Human studies have suggested that deficits in prefrontal cortical function and consequential loss of inhibitory control^{2–4} could be crucial in promoting compulsive drug use. However, it remains unknown whether chronic drug use compromises cortical activity and, equally important, whether this deficit promotes compulsive cocaine seeking. Here we use a rat model of compulsive drug seeking^{5–8} in which cocaine seeking persists in a subgroup of rats despite delivery of noxious foot shocks. We show that prolonged cocaine self-administration decreases *ex vivo* intrinsic excitability of deep-layer pyramidal neurons in the prelimbic cortex, which was significantly more pronounced in compulsive drug-seeking animals. Furthermore, compensating for hypoactive prelimbic cortex neurons with *in vivo* optogenetic prelimbic cortex stimulation significantly prevented compulsive cocaine seeking, whereas optogenetic prelimbic cortex inhibition significantly increased compulsive cocaine seeking. Our results show a marked reduction in prelimbic cortex excitability in compulsive cocaine-seeking rats, and that *in vivo* optogenetic prelimbic cortex stimulation decreased compulsive drug-seeking behaviours. Thus, targeted stimulation of the prefrontal cortex could serve as a promising therapy for treating compulsive drug use.

To investigate the neurobiological dysfunctions that underlie compulsive drug use, we used an animal model of addiction⁵ previously shown to engender a subpopulation of rats with compulsive cocaine seeking. Rats underwent extended cocaine self-administration training (Fig. 1a) under a heterogeneous seeking-taking chain schedule⁵ (details in the Supplementary Information). Daily self-administration sessions consisted of 30 trials that began with insertion of the seek lever. After one lever press and a variable interval (5 to 75 s), one additional press on the seek lever resulted in seek lever retraction and take lever insertion. One take lever press triggered a single intravenous cocaine infusion and retraction of the take lever. After >8 weeks of self-administration training, rats received four days of shock sessions in which 30% of the seek link ended with delivery of a noxious foot shock (0.4 mA, 0.5 s). Rats earning ≥ 10 cocaine infusions (14 of 46 rats, 30%, 24.1 ± 1.9 infusions) in the fourth shock session were considered shock-resistant, whereas rats earning <10 infusions (32 of 46 rats, 70%, 1.3 ± 0.3 infusions) were considered shock-sensitive (Supplementary Fig. 1). Previous works have observed a relatively similar proportion of shock-resistant rats^{5–8} and suggested that these rats exhibit greater basal impulsivity but not higher reactivity to novelty⁶. Seek lever presses (Fig. 1b, c) and earned cocaine infusions (Fig. 1d, e) were significantly inhibited by foot shocks in shock-sensitive rats, whereas shock-resistant rats were much less affected (Fig. 1b–e). Baseline cocaine-seeking behaviours and earned cocaine infusions were similar in both groups (Fig. 1c, e), suggesting that divergent responses to contingent foot shocks did not reflect cocaine self-administration history⁸.

Clinical evidence implicates prefrontal cortex (PFC) hypofunction^{2–4} in the loss of inhibitory control over drug seeking, but a causal link between drug-induced neuroadaptations and compulsive seeking behaviour remains elusive. We focused on the prelimbic cortex, because a homologous region in the human PFC^{9,10} regulates many cognitive functions, including decision making and inhibitory response control^{11–15}, which are compromised in human addicts³. We propose that chronic cocaine use induces prelimbic cortex hypoactivity, and that compromised prelimbic cortex function in turn impairs inhibitory

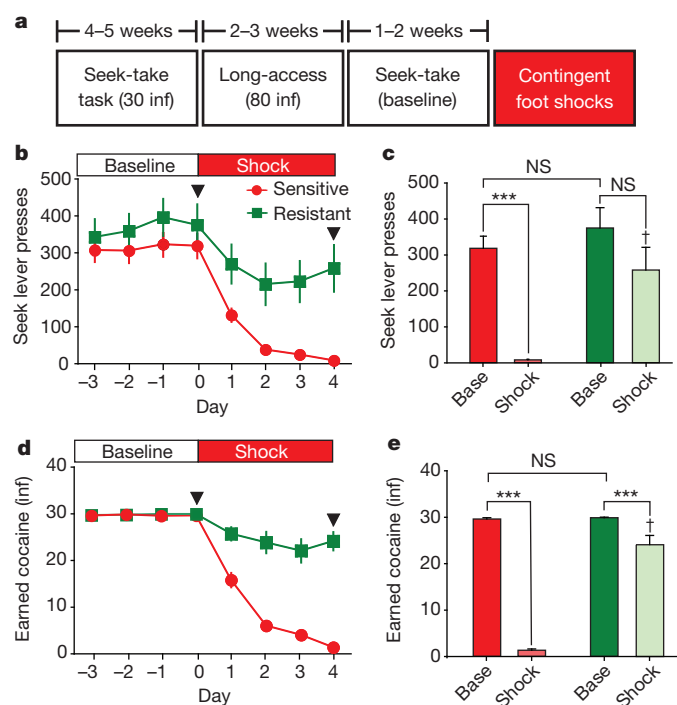


Figure 1 | Prolonged cocaine self-administration induces compulsive cocaine seeking in a subpopulation of rats. **a**, Experimental timeline showing the seek-take task and long-access sessions before the beginning of the foot shock sessions. After approximately two months of cocaine self-administration, rats received four days of shock sessions, in which a foot shock was administered in 30% of trials. **b**, Foot shocks decreased seek-lever presses in shock-sensitive rats ($n = 32$) but not shock-resistant rats ($n = 14$). **c**, Seek-lever presses on last day of baseline and last day of shock (arrowheads in **b**) ($F_{(3,68)} = 27.89$, $P < 0.001$; Tukey's post-hoc: *** $P < 0.001$ baseline versus shock sessions; † $P < 0.001$ shock-sensitive versus shock-resistant shock session). **d**, Foot shocks greatly reduced cocaine infusions in shock-sensitive rats. **e**, Earned cocaine infusions (inf) on last day of baseline and last day of shock ($F_{(3,68)} = 246.8$, $P < 0.001$; Tukey's post-hoc: *** $P < 0.001$ baseline versus shock session; † $P < 0.001$ shock-sensitive versus shock-resistant shock sessions). NS, not significant. Error bars indicate s.e.m.

¹Intramural Research Program, National Institute on Drug Abuse, Baltimore, Maryland 21224, USA. ²Ernest Gallo Clinic and Research Center, Department of Neurology, University of California San Francisco, San Francisco, California 94608, USA. ³Department of Neurology, University of California San Francisco, California 94110, USA. ⁴Solomon H. Snyder Neuroscience Institute, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA.

control over compulsive drug seeking. Thus, we performed *ex vivo* electrophysiology to determine whether cocaine self-administration reduced prelimbic cortex intrinsic excitability, because action potentials are the main mechanism of neuronal communication and altered spike firing contributes to neuronal dysfunction¹⁶ and central nervous system disorders¹⁷.

Intrinsic excitability of deep-layer, morphologically confirmed (Fig. 2a)¹⁸ pyramidal prelimbic cortex neurons from shock-sensitive and shock-resistant rats was evaluated 24 h after the fourth shock session, and was compared with naive rats. We targeted deep-layer pyramidal prelimbic cortex neurons because they project to brain structures implicated in drug-seeking behaviours, including the nucleus accumbens, dorsal striatum and amygdala^{15,19,20}. We first measured the amount of current required to induce an action potential in response to brief depolarization (2 ms, 0–2,500 pA, 10 pA steps). Relative to naive and shock-sensitive neurons, shock-resistant neurons required almost twice as much current to elicit an action potential

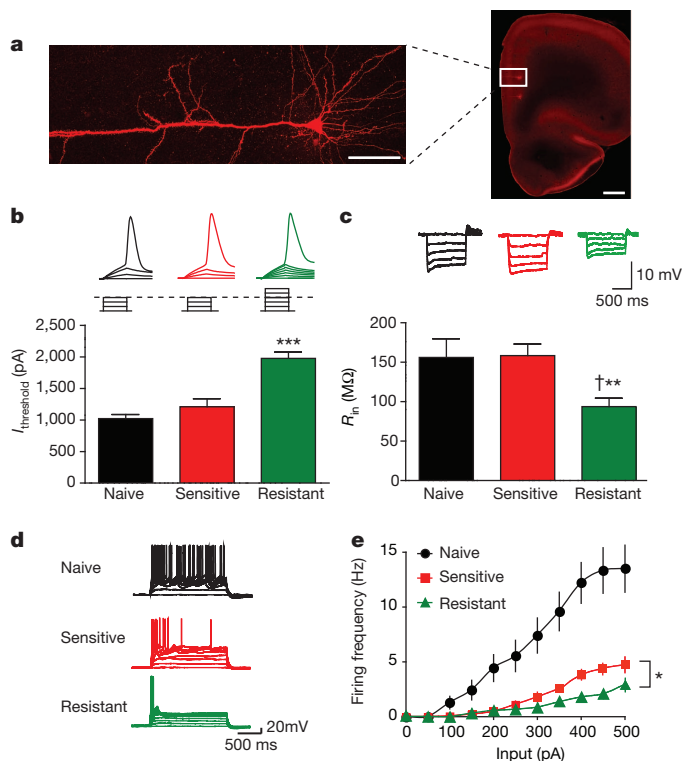


Figure 2 | Profoundly hypoactive prelimbic cortex neurons in shock-resistant rats. **a**, Morphological features of a recorded prelimbic cortex pyramidal neuron. Scale bars represent 0.1 mm and 1 mm in the left and right images, respectively. **b**, Shock-resistant prelimbic cortex neurons required more current to elicit firing with a 2-ms current step. Top, examples of sub-threshold membrane depolarization and the first action potential elicited by current injection. Square pulses illustrate increasing current injection. Bottom, current required for firing was higher in shock-resistant rats ($F_{(3,45)} = 18.53$, $P < 0.001$; Tukey's post-hoc: *** $P < 0.001$ naive or shock-sensitive versus shock-resistant) (naive, $n = 10$ cells; shock-sensitive, $n = 13$ cells; shock-resistant, $n = 16$ cells). **c**, Top, neuronal responses to hyperpolarizing steps. Bottom, input resistance was decreased only in shock-resistant rats ($F_{(3,60)} = 3.986$, $P = 0.011$; Tukey's post-hoc: † $P < 0.01$ naive versus shock-resistant, ** $P < 0.01$ shock-sensitive versus shock-resistant) (naive, $n = 12$ cells; shock-sensitive, $n = 24$ cells; shock-resistant, $n = 16$ cells). **d**, Prelimbic cortex firing in response to current injection. **e**, Firing frequency versus current injection (1 s duration, 0 to 500 pA, 50 pA steps) (two-way analysis of variance (ANOVA) among all groups: treatment, $F_{(3,59)} = 17.45$, $P < 0.001$; interaction, $F_{(30,590)} = 11.95$, $P < 0.001$; two-way ANOVA between shock-resistant and shock-sensitive: treatment, $F_{(1,33)} = 6.440$, * $P = 0.016$; interaction group \times input, $F_{(10,330)} = 6.201$, $P < 0.001$) (naive, $n = 15$ cells; shock-sensitive, $n = 19$ cells; shock-resistant, $n = 16$ cells). Statistical analyses include neurons from rats in Supplementary Fig. 5. Error bars indicate s.e.m.

(Fig. 2b) and exhibited significantly smaller input resistance (Fig. 2c). Thus, the ability of prelimbic cortex neurons to rapidly generate an action potential was compromised in shock-resistant rats. In addition, prelimbic cortex neurons from shock-resistant and shock-sensitive rats showed impairments in repetitive firing induced by a more moderate, sustained depolarizing currents (1 s, 0–500 pA, 50 pA steps), with a significantly greater impairment in shock-resistant versus shock-sensitive rats (Fig. 2d, e). Also, cocaine intake under shock correlated with decreased input resistance and increased current required for firing across all rats tested (Supplementary Fig. 2). Analysis of the action potential shape during whole-cell recordings and nucleated patch electrophysiological recordings revealed no differences in Na^+ channel activation, or in the composition of somatic K^+ currents or barium-sensitive K^+ conductances (Supplementary Table 1 and Supplementary Figs 3 and 4) among naive, shock-sensitive and shock-resistant rats. Control experiments in which shock-sensitive rats self-administered additional cocaine after each shock session suggested that the quantity of cocaine per se was not responsible for firing adaptations in shock-resistant rats (Supplementary Fig. 5).

Thus, our results demonstrate that long-term cocaine self-administration reduced prelimbic cortex excitability, with a much more pronounced effect in compulsive rats. If compromised spike firing was causally related to compulsive cocaine seeking, then increasing prelimbic cortex neuronal activity could decrease compulsive behaviour. One way to compensate for reduced excitability is to take advantage of *in vivo* optogenetic stimulation^{21,22}, in which prelimbic cortex neurons can be activated in a temporally precise manner during cocaine self-administration trials. Deep-layer prelimbic cortex neurons expressing channelrhodopsin-2 (ChR2) (Supplementary Fig. 6a) showed minimal loss of spike fidelity at photostimulation frequencies up to 20 Hz both *ex vivo* and *in vivo* (Supplementary Fig. 6b, c). Also, 1 Hz prelimbic cortex photoactivation evoked field potentials in the nucleus accumbens core, a downstream prelimbic cortex target²³, without loss of fidelity or generation of long-term depression (LTD) of glutamatergic function (Supplementary Fig. 7), although other frequencies can induce prelimbic cortex LTD²⁴. Thus, photoactivation of ChR2-expressing prelimbic cortex neurons can be reliably achieved at physiologically relevant timescales.

To test whether enhancing activity in a hypoactive prelimbic cortex would decrease compulsive cocaine seeking, adeno-associated virus (AAV) encoding ChR2 fused to enhanced yellow fluorescent protein (ChR2-eYFP) was bilaterally injected into the prelimbic cortex, with bilateral implantation of chronic optic fibres²² targeting the prelimbic cortex (Supplementary Fig. 8). The effect of prelimbic cortex stimulation on cocaine seeking was assessed in two separate sessions (Fig. 3A), and was only determined in the compulsive cocaine-seeking shock-resistant rats. The first prelimbic cortex ChR2 stimulation was administered during the last baseline session to determine the effect on control cocaine seeking. The second ChR2 stimulation session was given 24 h after the fourth shock session to determine whether prelimbic cortex activation could alter compulsive cocaine seeking; this session was identical to shock sessions except with the addition of photostimulation. In both ChR2 stimulation sessions, 1 Hz light pulses (10 ms duration, 473 nm) were delivered to the prelimbic cortex starting at the seek lever extension and terminating at the conclusion of the seek link. Prelimbic cortex stimulation had no effect on control, baseline cocaine-seeking behaviours (Fig. 3B, a–e). Similarly, *in vivo* prelimbic cortex photoactivation did not promote motivated behavioural responding (Supplementary Fig. 9).

In stark contrast to the baseline ChR2 stimulation session, prelimbic cortex photoactivation by ChR2 significantly attenuated foot-shock-resistant cocaine seeking. Photostimulation decreased earned cocaine infusions and increased the latency to seek-lever press (Fig. 3C, a and b), with a trend for decreased seek-lever presses and increased inter-press interval (Fig. 3C, c–e). Thus, restoring prelimbic cortex

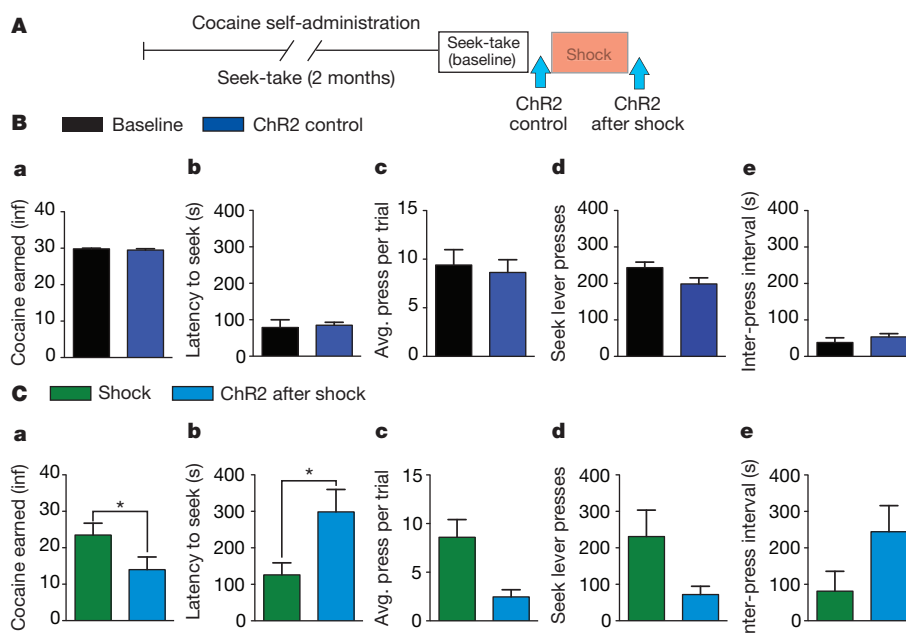


Figure 3 | In vivo optogenetic stimulation of prelimbic cortex suppresses compulsive cocaine seeking. **A**, Timeline illustrating when cocaine-seeking behaviours were analysed in shock-resistant rats. **B**, **C**, Comparisons were made between the last baseline session and ChR2 control session (**B**), and between the fourth shock session and the ChR2 plus shock session 24 h later (**C**). **B**, **a–e**, Prelimbic cortex photoactivation did not alter baseline cocaine self-administration. **C**, Light stimulation suppressed compulsive cocaine-seeking

behaviours, with decreased cocaine rewards earned (**a**, baseline/shock $F_{(1,5)} = 6.447$, $P = 0.052$; ChR2/no ChR2: $F_{(1,5)} = 13.998$, $P = 0.013$; interaction: $F_{(1,5)} = 17.699$, $P = 0.008$; Tukey's post-hoc: $*P < 0.05$) and increased latency to press the seek lever (**b**, baseline/shock $F_{(1,5)} = 5.756$, $P = 0.062$; ChR2/no ChR2: $F_{(1,5)} = 13.297$, $P = 0.015$; interaction: $F_{(1,5)} = 30.304$, $P = 0.003$; Tukey's post-hoc: $*P < 0.05$). A trend was observed in seek-lever presses (**c**, **d**) and inter-press interval (**e**). $n = 6$ rats. Error bars indicate s.e.m.

activity decreased compulsive cocaine seeking only after rats had experienced foot shocks. In addition, ChR2-mediated activation of prelimbic cortex reduced compulsive responding starting from the beginning of the cocaine-intake session (Supplementary Fig. 10),

suggesting that reduced intake may not require learning or other processes which might develop across the cocaine-intake session.

On the basis of these results, we proposed that prelimbic cortex inhibition might promote compulsive intake in shock-sensitive rats.

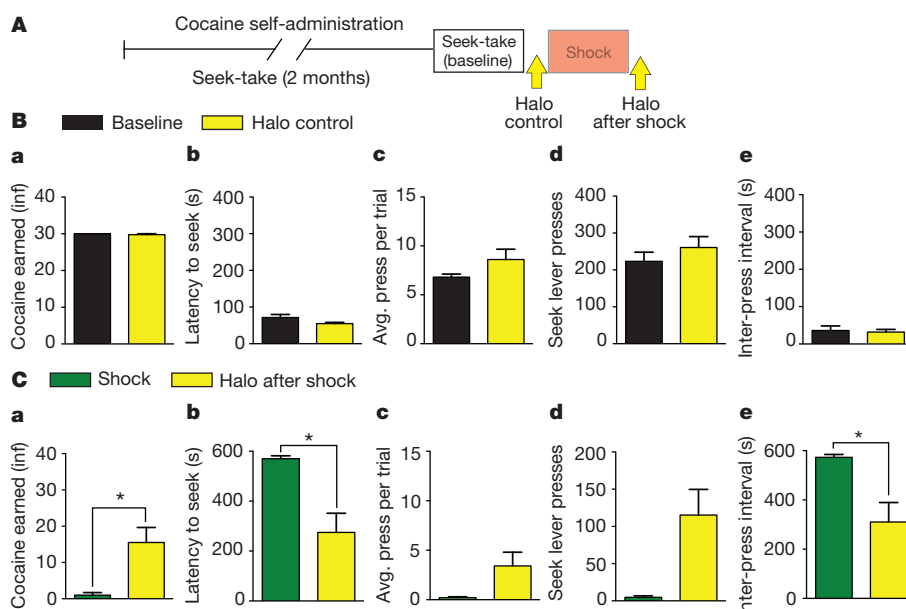


Figure 4 | In vivo optogenetic inhibition of prelimbic cortex enhances compulsive cocaine seeking. **A**, Timeline illustrating when cocaine-seeking behaviours were analysed. **B**, **C**, Comparisons were made between the last baseline session and eNpHR3.0 control session (**B**), and between the fourth shock session and the eNpHR3.0 plus shock session 24 h later (**C**). **B**, **a–e**, Prelimbic cortex photoactivation did not alter baseline cocaine self-administration. **C**, Light stimulation promoted cocaine-seeking behaviours in shock-sensitive rats, including increased cocaine rewards earned (**a**, baseline/shock $F_{(1,3)} = 126.99$, $P = 0.001$; eNpHR3.0/no eNpHR3.0: $F_{(1,3)} = 10.038$,

$P = 0.051$; interaction: $F_{(1,3)} = 12.271$, $P = 0.039$), decreased latency to press seek lever (**b**, baseline/shock $F_{(1,3)} = 118.59$, $P = 0.002$; eNpHR3.0/no eNpHR3.0: $F_{(1,3)} = 17.269$, $P = 0.025$; interaction: $F_{(1,3)} = 10.751$, $P = 0.046$) and decreased inter-press interval (**e**, baseline/shock $F_{(1,3)} = 124.12$, $P = 0.002$; eNpHR3.0/no eNpHR3.0: $F_{(1,3)} = 12.941$, $P = 0.037$; interaction: $F_{(1,3)} = 10.642$, $P = 0.047$). A trend was observed in seek-lever presses (**c**, **d**). $n = 4$ rats. NS, not significant. $*P < 0.05$ Tukey's post-hoc. Error bars indicate s.e.m. Halo, halorhodopsin (optogenetic inhibitor).

Thus, AAV encoding halorhodopsin-eYFP (eNpHR3.0-eYFP)^{21,22} was bilaterally injected into the prelimbic cortex, with bilateral chronic optic fibres targeting the prelimbic cortex (Supplementary Fig. 11). Prelimbic cortex inhibition by eNpHR3.0 was verified *ex vivo* (Supplementary Fig. 12). The effect of prelimbic cortex inhibition was assessed in two sessions with identical schedules as in the ChR2 experiments (Fig. 4A), except that neurons were inhibited by a continuous light pulse (532 nm) during the seek part of the seek-take chain. Prelimbic cortex inhibition had no effect on control, baseline cocaine-seeking behaviours (Fig. 4B, a–e), but significantly increased cocaine seeking during foot shock sessions in these shock-sensitive rats, with greater earned cocaine infusions (Fig. 4C, a), decreased latency to seek-lever press (Fig. 4C, b), decreased inter-press interval (Fig. 4C, e) and a trend for increased seek-lever presses (Fig. 4C, c and d). Thus, eNpHR3.0 inhibition of prelimbic cortex neurons increased propensity for compulsive cocaine seeking.

Our results suggest that prolonged cocaine use depressed prelimbic cortex excitability, and that, in a select population of rats, a profound prelimbic cortex hypoactivity drove compulsive cocaine seeking. We speculate that crossing a critical threshold of prelimbic cortex hypoactivity promotes compulsive behaviours. Notably, prelimbic cortex-mediated inhibitory control was only evident after self-administration was paired with foot shock, but not during baseline sessions before foot shock. In agreement, other groups have proposed that cortical areas promote compulsive intake in the presence of a challenge, for example, in the face of aversive consequences, in contrast to striatal areas that would sustain habitual intake in the absence of such challenges^{2,25}. Thus, in the absence of conflict, optogenetic manipulation of a brain region that mediates inhibitory control would have no effects. However, when confronted with the possibility of punishment, prelimbic cortex hypoactivity would bias behaviour towards compulsive seeking. Our results also suggest that prelimbic cortex stimulation might be clinically efficacious against compulsive seeking, with few side effects on non-compulsive reward-related behaviours in addicts; whether prelimbic cortex ChR2 stimulation reduces cocaine seeking through effects on emotional conditioning or cognitive processing^{11–15} remains to be resolved. In conclusion, we provide an effective method to decrease compulsive drug seeking in an animal model that may closely mimic the human addiction condition, which supports the use of *in vivo* prefrontal cortex stimulation to treat compulsive drug use.

METHODS SUMMARY

Cocaine self-administration. Outbred male Wistar (Harlan) rats weighing 275–300 g at the beginning of the experiments were used for all experiments. Cocaine self-administration began 7 days after surgery. Rats were trained 6 or 7 days per week during the animal's dark cycle.

Ex vivo patch-clamp recording. Rats were deeply anaesthetized, perfused with ice-cold sucrose-substituted buffer, decapitated, and 250- μ m sections of the medial PFC were prepared. Whole-cell current-clamp recordings and nucleated patch-clamp recordings were performed using a potassium-based internal solution^{26,27}.

In vivo optogenetic prelimbic cortex stimulation or inhibition during behaviour. Adeno-associated viral vectors (AAV5 CaMKII α :ChR2-eYFP or AAV5 CaMKII α :eNpHR3.0-eYFP) were packaged by the University of North Carolina Vector Core Facility. Following surgery to implant a jugular catheter for intravenous cocaine self-administration (see Methods), virus (0.5 μ l) was injected bilaterally into the prelimbic cortex at 0.1 μ l min⁻¹. Chronic optical fibres were implanted bilaterally directly above the prelimbic cortex. During optogenetic experiments, chronic optical implants were connected to optical patch cables coupled to 473-nm (for ChR2) or 532-nm (for eNpHR3.0) lasers modulated by behavioural hardware (Med Associates). Chronic optical implants were constructed on site as previously described²².

Full Methods and any associated references are available in the online version of the paper.

Received 1 March 2012; accepted 18 February 2013.

Published online 3 April 2013.

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn (2000).

2. Naqvi, N. H. & Bechara, A. The insula and drug addiction: an interoceptive view of pleasure, urges, and decision-making. *Brain Struct. Funct.* **214**, 435–450 (2010).
3. Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Rev. Neurosci.* **12**, 652–669 (2011).
4. Jentsch, J. D. & Taylor, J. R. Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacol.* **146**, 373–390 (1999).
5. Pelloux, Y., Everitt, B. J. & Dickinson, A. Compulsive drug seeking by rats under punishment: effects of drug taking history. *Psychopharmacol.* **194**, 127–137 (2007).
6. Belin, D., Mar, A. C., Dalley, J. W., Robbins, T. W. & Everitt, B. J. High impulsivity predicts the switch to compulsive cocaine-taking. *Science* **320**, 1352–1355 (2008).
7. Vanderschuren, L. J. & Everitt, B. J. Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science* **305**, 1017–1019 (2004).
8. Deroche-Gamonet, V., Belin, D. & Piazza, P. V. Evidence for addiction-like behavior in the rat. *Science* **305**, 1014–1017 (2004).
9. Uylings, H. B. M., Groenewegen, J. J. & Kolb, B. Do rats have a prefrontal cortex? *Behav. Brain Res.* **146**, 3–17 (2003).
10. Farvick, A., Dupont, L. M., Arce, M. & Eichenbaum, H. Medial prefrontal cortex supports recollection, but not familiarity, in the rat. *J. Neurosci.* **28**, 13428–13434 (2008).
11. Grégoire, S., Rivalan, M., Le Moine, C. & Dellu-Hagedorn, F. The synergy of working memory and inhibitory control: behavioral, pharmacological and neural functional evidences. *Neurobiol. Learn. Mem.* **97**, 202–212 (2012).
12. Hare, T. A., Camerer, C. F. & Rangel, A. Self-control in decision-making involves modulation of the vmPFC valuations. *Science* **324**, 646–648 (2009).
13. Balleine, B. W. & O'Doherty, J. P. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacol.* **35**, 48–69 (2009).
14. Jonkman, S., Mar, A. C., Dickinson, A., Robbins, T. W. & Everitt, B. J. The rat prelimbic cortex mediates inhibitory response control but not the consolidation of instrumental learning. *Behav. Neurosci.* **123**, 875–885 (2009).
15. Peters, J., Kalivas, P. W. & Quirk, G. J. Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn. Mem.* **16**, 279–288 (2009).
16. Krishnan, V. et al. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131**, 391–404 (2007).
17. Beck, H. & Yaari, Y. Plasticity of intrinsic neuronal properties in CNS disorders. *Nature Rev. Neurosci.* **9**, 357–369 (2008).
18. Yang, C. R., Seamans, J. K. & Gorelova, N. Electrophysiological and morphological properties of layers V–VI principal pyramidal cells in rat prefrontal cortex *in vitro*. *J. Neurosci.* **16**, 1904–1921 (1996).
19. Ghazizadeh, A., Ambroggi, F., Odean, N. & Fields, H. L. Prefrontal cortex mediates extinction of responding by two distinct neural mechanisms in accumbens shell. *J. Neurosci.* **32**, 726–737 (2012).
20. Sesack, S. R., Deutch, A. Y., Roth, R. H. & Bunney, B. S. Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with Phaseolus vulgaris leucoagglutinin. *J. Comp. Neurol.* **290**, 213–242 (1989).
21. Zhang, F., Wang, L.-P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
22. Witten, I. B. et al. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* **72**, 721–733 (2011).
23. Moussawi, K. et al. N-Acetylcysteine reverses cocaine-induced metaplasticity. *Nature Neurosci.* **12**, 182–189 (2009).
24. Kasanetz, F. et al. Prefrontal synaptic markers of cocaine addiction-like behavior in rats. *Mol. Psychiatry* <http://dx.doi.org/10.1038/mp.2012.59> (15 May 2012).
25. Tiffany, S. T. & Conklin, C. A. A cognitive processing model of alcohol craving and compulsive alcohol use. *Addiction* **95** (suppl. 2), S145–S153 (2000).
26. Martina, M., Schultz, J. H., Ehmk, H., Monyer, H. & Jonas, P. Functional and molecular differences between voltage-gated K⁺ channels of fast-spiking interneurons and pyramidal neurons of rat hippocampus. *J. Neurosci.* **18**, 8111–8125 (1998).
27. Taverna, S., Tkatch, T., Metz, A. E. & Martina, M. Differential expression of TASK channels between horizontal interneurons and pyramidal cells of rat hippocampus. *J. Neurosci.* **25**, 9162–9170 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Kourrich and Y. Shaham for careful reading of the manuscript. We also thank K. Deisseroth for providing the ChR2 and eNpHR3.0 vectors. This study was supported by funds from the NIDA/IRP and the State of California through the University of California at San Francisco.

Author Contributions B.T.C., H.-J.Y., F.W.H. and A.B. designed, discussed and planned all experiments. B.T.C., H.-J.Y., I.K.-Y., C.H. and S.L.C. performed experiments. B.T.C. and C.H. analysed data. B.T.C., F.W.H. and A.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.T.C. (billy.chen@nih.gov) or A.B. (antonello.bonci@nih.gov).

METHODS

Subjects. Adult male Wistar rats were singly housed in a climate-controlled facility. Experiments followed the guidelines of the National Institutes of Health Guide For the Care and Use of Laboratory Animals, and were performed during the animals' dark cycle.

Catheter surgery. Rats were implanted with a chronically indwelling intravenous catheter, as previously described²⁸. Briefly, rats were anaesthetized with isoflurane (2.5%) and a unilateral, intrajugular catheter was inserted 2.7–3.0 cm into the jugular vein. Catheters were sutured to the underlying muscle tissue and the other end externalized between the scapulas. Catheters consisted of a guide cannula (Plastics-One) attached to MicroRenathane tubing (0.025-inch inside diameter, 0.047-inch outside diameter) and dental cement. The catheter was flushed daily with a saline solution containing gentamicin (5 mg ml⁻¹, Butler Schein) to help maintain catheter patency and reduce infection. Following surgery, animals were allowed to recover for one week. During the first 5 days of the recovery period, animals were housed in home cages and had ad libitum access to food and water. Beginning on day 6, rats were food restricted to 20 g per day of standard rat chow. Rats were monitored daily to ensure that weight did not fall below 90% of ad libitum weight. Food restriction remained until the end of the experiment. Cocaine self-administration training began 7 days after surgery. Rats were food restricted to promote self-administration behaviour²⁹.

Stereotaxic surgery. Rats designated for optogenetic experiments received virus injection and chronic optic fibre surgery on the same day as catheter surgery. Rats were placed into a stereotaxic frame (David Kopf Instruments) and remained under isoflurane anaesthesia. Microinjection needles were then inserted bilaterally targeting the prelimbic area (prelimbic cortex; coordinates from Bregma: +3.2 AP, \pm 1.2 ML, -3.7 DV, 10° angle). Microinjections were performed using custom-made injection needles (26-gauge) connected to a 2- μ l Hamilton syringe. Each prelimbic cortex hemisphere was injected with 0.5 μ l of purified and concentrated AAV ($\sim 10^{12}$ infectious units ml⁻¹) encoding ChR2-eYFP or eNpHR3.0-eYFP under the control of the *Camk2a* promoter (AAV5 CaMKII α ::ChR2-eYFP or AAV5 CaMKII α ::eNpHR3.0-eYFP) to infect prelimbic cortex glutamatergic neurons. Injections occurred over 10 min followed by an additional 10 min to allow diffusion of viral particles away from the injection site. Chronic optic fibres were implanted bilaterally over the injection site (+3.2 anteroposterior, \pm 1.2 mediolateral, -3.4 dorsoventral, 10° angle). Fibres were secured to the skull with dental cement.

General behavioural methods. Behavioural experiments were performed in standard operant chambers (Med Associates or Coulbourn Instruments) with two retractable levers set 5 cm above the floor. A cue light was positioned 6 cm above each lever. The two levers flanked a food hopper. One lever was paired with cocaine infusion (the take lever) while the other lever was designated as the seek lever. Each chamber was also equipped with a house light and tone-generator. All sessions continued for 6 h or until 30 or 80 cocaine infusions were earned (see below for detail methods on seek-take chain schedule), and were performed 6 or 7 days a week for approximately 2 months.

Acquisition of cocaine self-administration under the seek-take schedule. Seven days after surgery, rats began cocaine self-administration training using the seek-take chain schedule. The behavioural model used was modified from the procedures described previously⁵. Self-administration training was divided into four separate phases: acquisition of the taking response; training of the seek-take chain; extended training; and punishment.

Acquisition of the taking response. In this initial phase of training, each trial began with the insertion of the take lever. Depression of the take lever on a fixed ratio (FR-1) resulted in the delivery of a single cocaine infusion (0.50 mg kg⁻¹ over 3 s). Cocaine infusions were paired with illumination of the cue light (20 s) above the take lever, retraction of the take lever and extinction of the house light. Following a 30 s time-out period, another trial was initiated with the insertion of the take lever and illumination of the house light. Training sessions terminated after either 30 earned cocaine infusions or 6 h, depending on which criterion was reached first. Training of the taking response continued until rats were able to earn 30 cocaine infusions across three consecutive sessions. Typically, training continued for five to seven sessions, after which rats advanced to training for the seek-take chain schedule.

Training of the seek-take chain. Each trial in the seek-take chained schedule began with insertion of the seek lever (opposite to the take lever), with the take lever retracted. A single press on the seek lever resulted in the retraction of the seek lever and insertion of the take lever. A single depression of the take lever then triggered delivery of a single cocaine infusion, illumination of the cue light above the take lever, retraction of the take lever and extinction of the house light. Following a 30 s time-out, another trial was initiated with the insertion of the seek lever and house light illumination.

Once the rats were able to earn 30 cocaine infusions, or had less than 20% variation in number of cocaine earned across three consecutive days, a random

interval (RI) schedule was introduced into the seek link of the chain schedule. Under the RI schedule, a trial began with the insertion of the seek lever and illumination of the house light. The first seek-lever press initiated the RI schedule. Three RI schedules were used across the training period: RI5, RI30 and RI60. In the RI5 schedule, immediately after the first seek-lever press, the software randomly selected 0.1 s, 5 s, or 10 s as the 'random interval'. Seek-lever presses during the duration of the RI had no programmed consequences. The first seek-lever press following the end of the RI (in the case of RI5: 0.1 s, 5 s, or 10 s) terminated the seek link of the chain and resulted in the retraction of the seek lever and the insertion of the take lever. In RI30 schedule, the software could randomly select 15, 30, or 45 s, whereas for RI60 intervals of 45 s, 60 s or 75 s were used. One press on the take lever triggered cocaine infusion, paired with the identical cues as during the training of the take response. A time-out period followed each cocaine infusion. The time-out period started at 30 s but progressively increased to 7 min by the end of training. The RI schedule was not altered within any individual training session. Training on each RI schedule persisted until the rats were able to earn 30 cocaine infusions, or have less than 20% variation across three consecutive days. Once an animal achieved stable cocaine seeking behaviour at a RI schedule, usually 4–6 days, it was advanced to the next RI schedule. Typically, a rat would spend 5 days in RI5, receive another 5 days in RI30, and finally finish with another 5 days at RI60. Rats were given 10 min to complete each seek-take chain. Failure to complete the cycle within the 10 min resulted in a lost opportunity to self-administer cocaine. A new trial started after the time-out period. More than 90% of rats in the present study earned 30 infusions under all three RI schedules. Daily training sessions terminated after either 30 earned cocaine infusions or 6 h, depending upon which criterion was reached first.

Extended training. After reaching the criterion for RI60, rats were given extended access to cocaine. In this phase of training, rats were given eight sessions with only the take lever available. Rats could earn up to 80 cocaine infusions under an FR-1 schedule with a post-reinforcement time-out of 30 s. Interspersed in between the extended cocaine sessions were additional sessions of training on the seek-take chain schedule under RI60 schedule. At the conclusion of the extended training sessions, all rats received additional training on the seek-take chain schedule under RI60 schedule. Stable performance (earning all 30 cocaine infusions or >20% variation) across four consecutive days served as baseline for comparison for the effects of punishment and *in vivo* optogenetic stimulation.

Punishment. During each punishment session, mild foot shocks (0.4 mA, 0.5 s) were administered in 30% of the trials. Shocks were delivered after the first lever press on the seek lever that occurred after the RI60 period was finished, and before extension of the take lever. Each punishment session consisted of 30 trials, and rats were given access to the take lever for cocaine infusions in all trials. Rats were given four punishment sessions. At the end of four days, shock-sensitive and shock-resistant rats were grouped based on the number of earned cocaine rewards.

***In vivo* optical excitation sessions with ChR2.** Shock-resistant rats received *in vivo* optogenetic excitation of the prelimbic cortex in which light stimulation (1 Hz, 10 ms, 10–15 mW, 473 nm)^{21,22} was administered in two separate sessions. The first session was administered one day after last RI60 baseline sessions (Fig. 3a). A separate session was administered one day after the fourth shock sessions; thus, there was a fifth day of foot shock during which ChR2 was stimulated using a laser. In each of the two ChR2 stimulation sessions, light stimulation was administered only during the seek link of the seek-take chain schedule. Insertion of the seek lever initiated activation of the laser. A 1 Hz stimulation was present throughout the seek link. Photostimulation was terminated at the end of the seek link.

***In vivo* optical inhibition sessions with eNpHR3.0.** Shock-sensitive rats received *in vivo* optogenetic inhibition of the prelimbic cortex, in which light stimulation (continuous, 10–15 mW, 532 nm)^{21,22} was administered in two separate sessions. The first session was administered one day after the last RI60 baseline sessions (Fig. 4a). The second session was administered one day after the four shock sessions; thus, there was a fifth day of foot shock during which eNpHR3.0 was stimulated using a laser. Insertion of the seek lever initiated activation of the laser. Laser stimulation was continuously present throughout each seek link of the seek-take chain schedule. Photostimulation was terminated at the end of the seek link.

Drugs. Cocaine HCl, picrotoxin and 6-cyano-7-nitroquinoxaline-2,3-dione disodium salt hydrate (CNQX) were purchased from Sigma-Aldrich.

Slice preparation. Rats were euthanized one day after the last shock session. Rats were anaesthetized with 40 mg kg⁻¹ pentobarbital (intraperitoneally) and transcardially perfused with ~ 30 ml of nearly frozen ($\sim 0^\circ$ C) modified artificial cerebrospinal fluid (aCSF) at a rate of ~ 20 ml min⁻¹. The modified aCSF for perfusion contained (in mM): 225 sucrose, 119 NaCl, 2.5 KCl, 1.0 NaH₂PO₄, 4.9 MgCl₂, 0.1 CaCl₂, 26.2 NaHCO₃, 1.25 glucose. After perfusion, the brain was quickly removed and placed into ice-cold aCSF for 1–2 min. Coronal sections containing the medial PFC (250 μ m) were prepared with VT-1200 vibratome (Leica). Slices

were placed in a holding chamber (containing aCSF with 1 mM ascorbic acid added 15 min before brain dissection), and allowed to recover for at least 30 min before being placed in the recording chamber and superfused with a bicarbonate-buffered solution saturated with 95% O₂ and 5% CO₂ and containing (in mM): 119 NaCl, 2.5 KCl, 1.0 NaH₂PO₄, 1.3 MgCl₂, 2.4 CaCl₂, 26.2 NaHCO₃, and 11 glucose, at 32–34 °C.

Electrophysiology. Picrotoxin (100 µM) and CNQX (10 µM) were present throughout the experiment to block inhibitory and excitatory synaptic transmission, respectively. Cells were visualized using infrared differential interference contrast video microscopy. Whole-cell current-clamp recordings were made using a MultiClamp 700B amplifier (Molecular Devices). Electrodes (2.8–4.0 MΩ) contained (in mM): 120 potassium methanesulphonate (KMeSO₄), 20 HEPES, 0.4 EGTA, 2.8 NaCl, 5 TEA-Cl, 2.5 MgATP, and 0.25 NaGTP, biocytin (1%), pH 7.2–7.3 (270–285 mOsm). Series resistance (10–40 MΩ) was continually monitored on-line with a –20 pA, 300 ms current injection given after every current injection step; if the series resistance changed by more than 20%, data were not included in the analysis. Membrane potentials were not corrected for junction potentials (estimated to be 10 mV). After breaking into a neuron, the resting membrane potential was set to –80 mV by injecting DC current through patch amplifier. To measure the amount of current required to reach action potential threshold, a series of current steps (2 ms duration at 2.5 Hz, 0 to 2500 pA range with +10 pA step increments) were injected into the cell until an action potential was generated. To determine the input resistance, hyperpolarizing current injections (1 s duration, 0 to –100 pA in –25 pA step increments) were injected into the cell. Input resistance was taken at the linear part of the trace. Sustained firing was determined from a series of 11 current injections (1 s duration, 50 pA steps). All values were obtained after the cells had reached a stable response and were averages of three cycles for each cell.

Nucleated patch recordings were performed as described before to dissect K⁺ current components²⁶. Briefly, recording pipettes were filled with K⁺ based internal solution (in mM): 140 KMeSO₄, 5 KCl, 0.05 EGTA, 2 MgCl₂, 2 Na₂ATP, 0.4 NaGTP and 10 HEPES (pH adjusted to 7.3). Recordings were made at room temperature (23–24 °C). Immediately after establishing whole-cell configuration from identified pyramidal cells, a negative pressure (–150 to –190 mbar) was applied through patch pipette to bring the nucleus close to the pipette tip. The patch pipette was then slowly withdrawn until the membrane patch surrounding the nucleus was completely pulled out of the cell. After excision, nucleated patches were held with a small constant negative pressure (–30 to –60 mbar) and were voltage clamped at a holding potential of –90 mV. P/–4 protocol in pClamp 10 was used for on-line subtraction of leakage and capacitive currents. A custom-made multi-barrel system consisting three glass capillaries was used to apply drugs to isolate different voltage-gated K⁺ currents. Tetrodotoxin (TTX, 500 nM) and cadmium (Cd²⁺, 100 µM) were included in HEPES-buffered ACSF (in mM: 138 NaCl, 10 HEPES, 2.5 KCl, 2 CaCl₂, 1 MgCl₂ and 25 glucose, pH 7.3) in each barrel to block voltage-gated Na⁺ and Ca²⁺ currents, respectively. The K⁺ current blockers 4-aminopyridine (4-AP, 0.5 mM) and tetraethylammonium (TEA, 20 mM) were included in different barrels, and nucleated patches were inserted into the barrels sequentially. Macroscopic voltage-gated K⁺ current was elicited by a 150 ms test pulse stepping from –90 mV to +70 mV and up to 10 sweeps of recordings were

collected and averaged to represent the current in each drug treatment. Drug-sensitive K⁺ currents were then obtained by off-line subtraction and peak currents were measured to calculate the respective contribution in macroscopic K⁺ current. To study background K⁺ conductance, voltage-clamp recordings were made in the presence of 500 nM TTX, 0.1 mM Cd²⁺, 10 mM TEA and 1 mM CsCl to block voltage-gated Na⁺, Ca²⁺, K⁺ currents and I_h/inward rectifier currents, respectively. Cells were voltage-clamped at –70 mV. Under this condition, a slow voltage ramp from –130 mV to –70 mV was applied across 1 s before and after 1 mM barium (Ba²⁺) application. Pharmacological subtraction revealed Ba²⁺-sensitive current and background K⁺ conductance was then calculated at –70 mV according to ref. 27.

All recorded neurons were filled with biocytin and processed for post-hoc immunohistochemical staining to confirm the presence of apical dendrites. Only neurons exhibiting a clear apical dendrite, indicative of pyramidal neurons, were included in the analysis. All data were acquired at 20 kHz and filtered at 10 kHz using Clampex 10.2 software (Molecular Devices).

Data analysis. Statistical significance was primarily assessed using ANOVA, followed by post-hoc test when applicable. Data were analysed with Prism (GraphPad Software).

Virus expression and histology. Following the completion of behavioural experiments, rats were deeply anaesthetized with euthasol and perfused transcardially with PBS followed by 4% paraformaldehyde dissolved in PBS. Brains were removed and fixed in 4% paraformaldehyde for an additional 24–48 h. Fifty-micrometre-sections of the PFC were made on a vibratome. Some slices were stained for 1 h with 2% Neurotrace fluorescent Nissl stain (Invitrogen; excitation 530 nm, emission 615 nm) diluted in PBS. Slices were then washed and mounted on gelatin-coated slides, treated with fluorescent-mounting media and mounted. Some slices were not stained with Neurotrace and were directly washed with PBS and mounted. In these slices, 4',6-diamidino-2-phenylindole (DAPI) (0.15 µg ml^{–1}) was included in the fluorescent-mounting media. Expression of ChR2-eYFP and eNpHR3.0-eYFP was examined for all rats using a Nikon inverted fluorescent microscope with a ×4 objective or an Olympus Fluoview FV1000 with a ×4 objective. Rats showing no bilateral eYFP expression in the PFC owing to faulty microinjections, and showing bilateral fibre placement outside the PFC were excluded from analysis.

Reconstruction of optical stimulation sites in the prelimbic cortex. Optical stimulation sites of *in vivo* optogenetic experiments were determined as described previously³⁰. Briefly, fixed and stained coronal brain sections containing the mPFC were examined with Nikon inverted fluorescent microscope or an Olympus Fluoview FV1000. Optic fibre tracks were located in the slice and optical stimulation sites were determined as 0.5 mm below the fibre tip (Supplementary Figs 8, 9 and 11).

28. Chen, B. T. *et al.* Cocaine but not natural reward self-administration nor passive cocaine infusion produces persistent LTP in the VTA. *Neuron* **59**, 288–297 (2008).
29. Knackstedt, L. A. & Kalivas, P. W. Extended access to cocaine self-administration enhances drug-primed reinstatement but not behavioral sensitization. *J. Pharmacol. Exp. Ther.* **322**, 1103–1109 (2007).
30. Stuber, G. D. *et al.* Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475**, 377–380 (2011).

Visualization of an endogenous retinoic acid gradient across embryonic development

Satoshi Shimozono¹, Tadahiro Iimura¹, Tetsuya Kitaguchi², Shin-ichi Higashijima³ & Atsushi Miyawaki^{1,2}

In vertebrate development, the body plan is determined by primordial morphogen gradients that suffuse the embryo. Retinoic acid (RA) is an important morphogen involved in patterning the anterior–posterior axis of structures, including the hindbrain^{1–6} and paraxial mesoderm^{7,8}. RA diffuses over long distances, and its activity is spatially restricted by synthesizing and degrading enzymes⁹. However, gradients of endogenous morphogens in live embryos have not been directly observed; indeed, their existence, distribution and requirement for correct patterning remain controversial¹⁰. Here we report a family of genetically encoded indicators for RA that we have termed GEPRAs (genetically encoded probes for RA). Using the principle of fluorescence resonance energy transfer we engineered the ligand-binding domains of RA receptors to incorporate cyan-emitting and yellow-emitting fluorescent proteins as fluorescence resonance energy transfer donor and acceptor, respectively, for the reliable detection of ambient free RA. We created three GEPRAs with different affinities for RA, enabling the quantitative measurement of physiological RA concentrations. Live imaging of zebrafish embryos at the gastrula and somitogenesis stages revealed a linear concentration gradient of endogenous RA in a two-tailed source–sink arrangement across the embryo. Modelling of the observed linear RA gradient suggests that the rate of RA diffusion exceeds the spatiotemporal dynamics of embryogenesis, resulting in stability to perturbation. Furthermore, we used GEPRAs in combination with genetic and pharmacological perturbations to resolve competing hypotheses on the structure of the RA gradient during hindbrain formation and somitogenesis. Live imaging of endogenous concentration gradients across embryonic development will allow the precise assignment of molecular mechanisms to developmental dynamics and will accelerate the application of approaches based on morphogen gradients to tissue engineering and regenerative medicine.

RA is a small lipophilic molecule that acts as a ligand for nuclear RA receptors (RARs). RA is synthesized from retinal by retinaldehyde dehydrogenase type 2 (Raldh2) and degraded into polar metabolites by Cyp26 (Supplementary Fig. 1a). During early vertebrate embryogenesis, regions of RA synthesis and degradation were mapped by detecting messenger RNA encoding these metabolizing enzymes. In zebrafish, *raldh2* is expressed in the mid-trunk, whereas *cyp26* is expressed at both the anterior and posterior ends¹¹. We therefore speculate that a two-tailed gradient of RA could form with the highest concentration in the mid-trunk and tapering off at each end (Supplementary Fig. 2). Gradient formation of other morphogens, all of which are genetically encoded peptides, could be observed by fusing them to green fluorescent protein (GFP) and introducing them exogenously into embryos^{12–16}. However, the non-peptidic structure of RA makes it difficult to image directly¹⁷. It therefore remains unknown whether postulated RA gradients exist, have a linear distribution and are required for normal embryonic development.

To address this gap we developed GEPRAs. The ligand-binding domains (LBDs) from mouse RARs were flanked by cyan fluorescent

protein (CFP) and yellow fluorescent protein (YFP), mutants of *Aequorea* GFP (Fig. 1a). Alterations in the conformation of the LBD in response to RA binding are converted into changes in fluorescence resonance energy transfer (FRET) from CFP to YFP. Among the constructs containing the LBD from RAR- β (Supplementary Fig. 3a), the GEPRAs fusion protein showed the largest response in HeLa cells (Supplementary Fig. 4). *In situ* calibration for the intracellular RA concentration ($[RA]_i$) demonstrated an apparent dissociation constant (K'_d) of 2 nM (Fig. 1b). Responses to the RA precursors retinal and

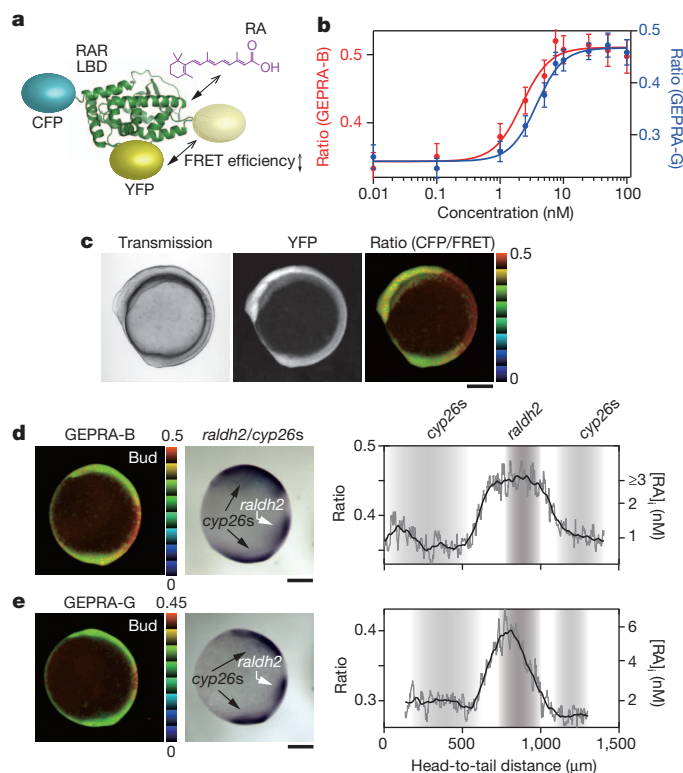


Figure 1 | Development and characterization of GEPRAs. **a**, Schematic representation of GEPRAs. The light and dark yellow denote the unbound and RA-bound states of GEPRAs, respectively. **b**, RA titration curves for GEPRAs-B (red) and GEPRAs-G (blue). R_{max} (RA-depleted form) and R_{min} (RA-saturated form) values for GEPRAs-B were slightly higher than those of GEPRAs-G. Fitting with the Michaelis–Menten equation yielded K'_d values of 2 and 4 nM for GEPRAs-B and GEPRAs-G, respectively. Each data point is the mean \pm s.d. for nine experiments. **c**, Transmission and fluorescence (YFP and FRET ratio) images of a five-somite embryo from the GEPRAs-B transgenic zebrafish line. **d**, **e**, Ratiometric $[RA]_i$ images (left), whole-mount *in situ* hybridization results for *raldh2* and *cyp26s* (centre), and graphs of the spatial distributions of *raldh2* and *cyp26s* (right) in bud-stage embryos expressing GEPRAs-B (**d**) and GEPRAs-G (**e**). Absolute $[RA]_i$ values are displayed on the right side of each graph. Scale bars, 200 μ m.

¹Laboratory for Cell Function Dynamics, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-city, Saitama, 351-0198, Japan. ²Life Function and Dynamics, ERATO, JST, 2-1 Hirosawa, Wako-city, Saitama, 351-0198, Japan. ³National Institutes of Natural Sciences, Okazaki Institute for Integrative Bioscience, National Institute for Physiological Sciences, Okazaki, Aichi, 444-8787, Japan.

retinol were almost negligible ($K'_d \gg 100$ nM; Supplementary Fig. 1b). We speculated that using multiple GEPRAs with different RA affinities would allow us to measure $[RA]_i$ quantitatively and create two additional GEPRAs. GEPRAs-A was generated by introducing two amino-acid substitutions in the LBD of GEPRAs-B. Of the constructs containing the LBD from RAR- γ , GEPRAs-G produced the best results (Supplementary Fig. 3b). GEPRAs-AA and GEPRAs-G produced K'_d values of 50 nM (Supplementary Fig. 5a) and 4 nM (Fig. 1b), respectively.

We generated transgenic zebrafish lines ubiquitously expressing GEPRAs-B or GEPRAs-G. In a five-somite embryo expressing GEPRAs-B, the probe was uniformly distributed (Fig. 1c, YFP), but the CFP/FRET ratio—that is, $[RA]_i$ —was highest in the mid-trunk and lowest in the head and tail (Fig. 1c, ratio). Next we depleted embryos of RA with 10 μ M 4-(diethylamino)benzaldehyde (DEAB), an inhibitor of Raldh2. DEAB abolished the GEPRAs-B-derived high $[RA]_i$ signal in the mid-trunk (Supplementary Fig. 6). This indicates that the probe is RA-specific because DEAB should increase retinal and retinol concentrations.

We compared signals representing $[RA]_i$ with the spatial expression patterns of RA-metabolizing enzymes. Immediately after imaging with GEPRAs-B (Fig. 1d, left), bud-stage embryos were fixed and subjected to *in situ* hybridization with probes specific for *raldh2* and *cyp26* species (*cyp26s*; a mixture of *cyp26a1*, *cyp26b1* and *cyp26c1*)⁵ (Fig. 1d, centre). Comparative spatial profiles indicated that the signals representing $[RA]_i$ were high in the *raldh2*-expressing region and low in the *cyp26s*-expressing region (Fig. 1d, right). Two intermediate zones flanked by the *raldh2*-expressing and *cyp26s*-expressing regions were of particular interest. In the posterior zone, the signal representing $[RA]_i$ was graded almost linearly. In the anterior zone, however, the probe saturated near the *raldh2*-expressing region. To examine high levels of $[RA]_i$ quantitatively, we imaged GEPRAs-G transgenic zebrafish at the bud stage (Fig. 1e) and found linear $[RA]_i$ gradients in both the anterior and posterior intermediate zones. We also expressed GEPRAs-AA transiently by injecting fertilized wild-type eggs with mRNA. GEPRAs-AA produced no $[RA]_i$ gradient in three-somite embryos (Supplementary Fig. 5b). On the basis of the K'_d values of the three GEPRAs, the highest $[RA]_i$ situated anteriorly within the *raldh2*-expressing region was estimated to be 6 nM (Fig. 1e). All of the gradients identified by using both GEPRAs-B and GEPRAs-G (Fig. 1d, e) reached the edges of the *cyp26s*-expressing regions, indicating that the steady-state distribution of RA is linear. This finding supports the model in which a local source and a local sink together generate a linear gradient in the flanked region based on a simple diffusion model¹⁸ within the spatiotemporal parameters of embryogenesis¹⁹. We examined whether a rectangular distribution of $[RA]_i$ could exist stably in the intermediate zones with a computer simulation. Assuming that the diffusion coefficient of RA and the width of the zone were $10 \mu\text{m}^2 \text{s}^{-1}$ (ref. 20) and 200 μm , respectively, our simulation showed that a rectangular distribution created a linear gradient in about 10 min (Supplementary Fig. 7), which is much faster than the time scale of embryonic growth and supports a source–sink model in the formation of a linear gradient by simple diffusion.

RA signalling in zebrafish embryos was previously observed using an exogenous reporter gene system, in which an RA response element (RARE) drives the expression of enhanced YFP (eYFP)²¹. However, RARE–eYFP transgenic zebrafish did not produce specific fluorescence signals until very late, at roughly the 18-somite stage (Supplementary Fig. 8a), and even at the 20-somite stage eYFP fluorescence was detected only in old somites. In contrast, the GEPRAs-B signal indicating a high $[RA]_i$ was distributed from the region near the otic vesicle to the most posterior somite (Supplementary Fig. 8b). At this stage, *raldh2* mRNA accumulated in the eyes and all somites, whereas mRNAs encoding *cyp26s* were abundant in the head region and tail tip (Supplementary Fig. 8c); these patterns agree with the distribution of GEPRAs-B signals representing $[RA]_i$. Thus, RARE–eYFP, which unlike GEPRAs is an indicator of RA signalling and not $[RA]_i$, is not

sensitive in zebrafish embryos because the eYFP chromophore takes a relatively long time to mature compared with the rapid timescale of zebrafish embryogenesis.

We verified that the $[RA]_i$ gradient in the head region at the bud stage (Fig. 1d, e) was contained within the hindbrain field (Supplementary Fig. 9). We also found that the signal representing high $[RA]_i$ in the mid-trunk was detectable at 75% epiboly and developed into a clear peak at the tailbud stage (Fig. 2a). Given these observations, we were able to resolve a historical controversy about the putative RA gradient in the hindbrain^{4–6}. The major argument against the presence of this gradient is the observation that embryos depleted of endogenous RA can be fully rescued with a uniform concentration of exogenous RA⁵. To examine whether this approach resulted in a rectangular RA distribution, we performed rescue experiments with the GEPRAs-B transgenic line (Fig. 2b). At 4 h after fertilization (hpf), embryos were exposed to 10 μ M DEAB and various concentrations of RA. The embryos were assessed for $[RA]_i$ at the three-somite to four-somite stages and for morphology at 36 hpf. Treatment with 10 μ M DEAB (Fig. 2b, second column of images) nearly abolished the signal representing high $[RA]_i$ and the imaged embryo developed a kinked head, which is a characteristic of RA depletion in zebrafish². When 10 nM RA was applied together with 10 μ M DEAB (Fig. 2b, fourth

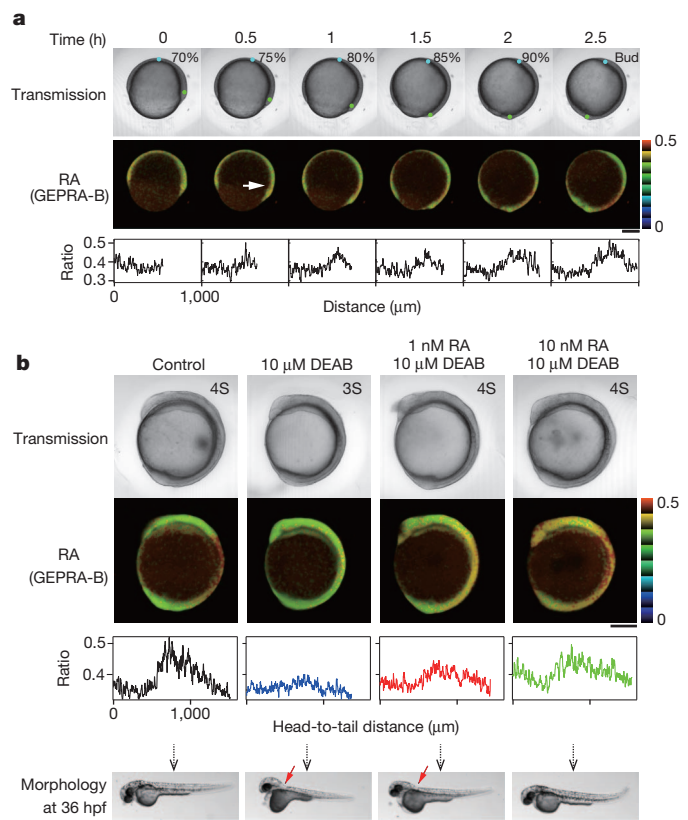


Figure 2 | $[RA]_i$ gradient during hindbrain development. **a**, Time-lapse $[RA]_i$ imaging of GEPRAs-B in an embryo from 70% epiboly to bud stage. A series of transmission images (top), $[RA]_i$ images (middle) and ratio profiles (bottom) are shown. The most anterior and most posterior points represented in the ratio profiles are indicated in the transmission images by cyan and green dots, respectively. Times since the start of imaging are shown above the figures. **b**, Visualization of $[RA]_i$ in GEPRAs-B-bearing embryos (three-somite (3S) to four-somite (4S) stage) with and without DEAB to deplete RA, and various concentrations of exogenous RA as denoted above the images. In each column reflecting the various treatment conditions, transmission and $[RA]_i$ images are accompanied by a graph showing the spatial distribution of $[RA]_i$ along the anterior–posterior axis of the body. A transmission image of a later developmental stage (36 hpf) is also presented. A red arrowhead indicates a kink in the head region. Scale bars, 200 μm .

column), however, a signal reflecting high $[RA]_i$ was detected. Although full recovery was not observed, a substantial $[RA]_i$ gradient was created in the hindbrain field. The gradient was sufficient to allow normal embryo development. Under the same conditions, 1 nM RA (Fig. 2b, third column) resulted in a smaller $[RA]_i$ gradient and did not rescue the zebrafish from the effects of RA depletion. The results were reproduced for each treatment condition with multiple embryos (Supplementary Fig. 10). These perturbation experiments revealed that the distribution of RA for hindbrain patterning is reliably governed by local degradation of RA, as proposed previously^{4–6}. Such $[RA]_i$ gradients depending on RA degradation were also observed during somitogenesis (Supplementary Fig. 11). In the zebrafish mutant giraffe (*gir*) containing a mutation in the *cyp26a1* gene, there are patterning defects in various organs²². We injected GEPRAs-B mRNA into *gir* embryos, and did indeed find a global increase in $[RA]_i$ in six-somite embryos (Supplementary Fig. 12).

Another essential morphogen that functions during the formation of the embryonic axis includes the fibroblast growth factor (FGF) family^{7,8,23,24}. The interactions between RA and FGF signalling include positive and negative feedback and feedforward mechanisms. To examine how Fgf8 regulates endogenous RA gradients, we performed $[RA]_i$ imaging experiments using embryos in which Fgf8 signalling was suppressed. We injected *fgf8* splice-blocking morpholino oligonucleotides (MOs)²⁵ into one-cell-stage embryos to phenocopy the well-characterized zebrafish mutant *acerebellar* (*ace*)^{26,27}. As a result of mutation of the *fgf8* gene, homozygous *ace* embryos lack both a cerebellum and organizer in the midbrain–hindbrain boundary while retaining expression of some rhombomere-marker genes²⁶. Because RA signalling is required for the formation of posterior segments of the hindbrain^{2–6}, such as rhombomeres 5–7, we assessed whether the RA gradient was retained in the hindbrain field of *fgf8* morphants expressing GEPRAs-B. Time-lapse transmission images confirmed effective knockdown of *fgf8* expression based on a bulge in the area of the developing midbrain (arrowheads)²⁶ and shortened tail (Fig. 3a, b). In the MO-treated embryos, the signal representing high $[RA]_i$ in the mid-trunk was observed until 12.5 hpf (Fig. 3b). Normal posteriorization of their hindbrain²⁶ was verified by using *in situ* hybridization with two rhombomere markers (Supplementary Fig. 13). In addition, the expression profiles of *raldh2* and *cyp26s* were unaffected at 10 hpf (Fig. 3c). After 13.5 hpf, however, high $[RA]_i$ was attenuated to baseline levels in *fgf8* morphants (Fig. 3b). We found that this decrease in $[RA]_i$ was due to decreased levels of *raldh2* expression (Fig. 3d), indicating that *fgf8* may have a role in the maintenance of *raldh2* gene expression. The downregulation of *raldh2* expression was previously observed in *ace* mutants²⁸; however, the GEPRAs system allowed us to directly image the effect on the RA gradient after the loss of *raldh2* expression.

The free diffusion of a lipophilic molecule such as RA may be limited in aqueous environments. A large fraction of RA molecules are probably transported intracellularly by interaction with cellular RA-binding proteins^{10,29}. On the assumption that most intracellular RAs are bound to carrier proteins, GEPRAs detect the free form of RA. The endogenous RA-buffering system is reminiscent of intracellular Ca^{2+} buffering in which loading cells with a large amount of high-affinity Ca^{2+} probes does not markedly affect intracellular Ca^{2+} dynamics³⁰. Thus, the expression of GEPRAs can be increased without disturbing endogenous RA dynamics or the normal embryonic development of fish. In support of this, GEPRAs transgenic embryos developed without apparent morphological aberrations.

Here we have generated fluorescent probes that allowed direct visualization of the endogenous RA gradient in live embryos. The existence of an RA gradient is a matter of historical debate^{4–6}. Our GEPRAs system clearly demonstrated that a two-tailed linear RA gradient exists during early embryogenesis, and that Fgf8 is not required for RA gradient formation in the hindbrain but is required for the maintenance of *raldh2* expression and the RA gradient later in development. The ability of GEPRAs to directly reveal endogenous morphogen

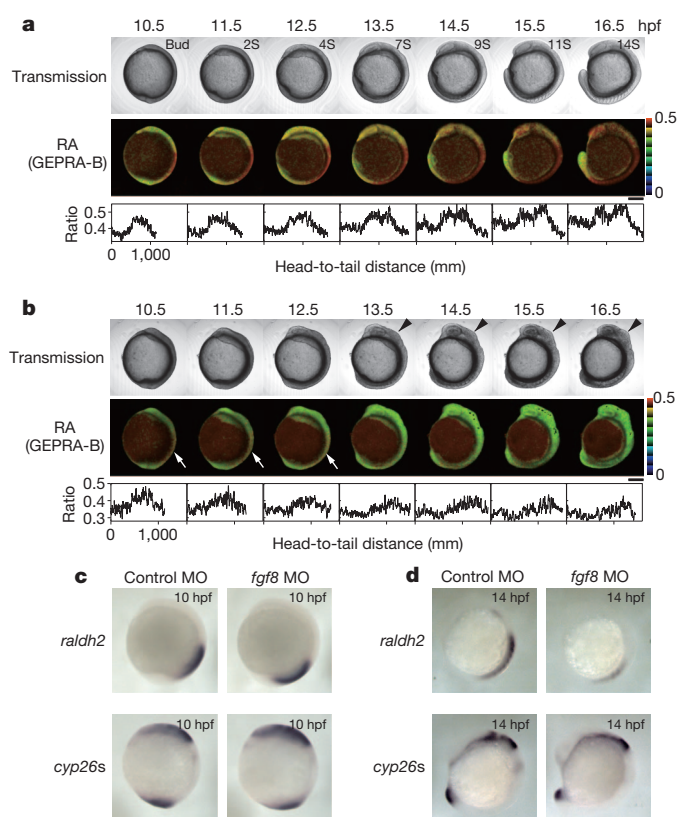


Figure 3 | $[RA]_i$ gradient is affected by *fgf8* expression. **a, b**, Time-lapse $[RA]_i$ imaging of GEPRAs-B-bearing embryos from 10.5 to 16.5 hpf after injections of control MO (**a**) or *fgf8*-specific MO (**b**). Transmission images (top row), $[RA]_i$ images (middle row) and ratio profiles (bottom row) are presented. The control embryo showed normal somitogenesis; somite stages are labelled in the transmission images (**a**). A bulge in the area of the developing midbrain in the embryo injected with the *fgf8*-specific MO is indicated by black arrowheads; a signal representing a high $[RA]_i$ in the trunk is indicated by white arrows (**b**). **c, d**, Comparative *in situ* hybridizations showing expression of *raldh2* and *cyp26s* in embryos at 10 hpf (**c**) and 14 hpf (**d**) after injections with control or *fgf8*-specific MO. All images are lateral views. Scale bars, 200 μ m.

gradients such as RA in live embryos will allow a greater understanding of their roles and mechanisms in patterning the vertebrate embryo and will support gradient-based approaches in medicine and bioengineering.

METHODS SUMMARY

The genes for GEPRAs were constructed from complementary DNAs encoding the LBD of mouse RAR- β or RAR- γ , and subcloned into pCS2 for expression in HeLa cells and for *in vitro* synthesis of mRNA or into pT2KXIG Δ in for generation of transgenic zebrafish lines. $[RA]_i$ imaging was performed with an inverted confocal microscope equipped with a $\times 10$ objective lens and a 440-nm laser. With the aperture fully open, non-confocal fluorescence images were acquired. The dual-emission ratio imaging was performed under exactly the same conditions for *in situ* $[RA]_i$ calibration (HeLa cells) and *in vivo* time-lapse two-dimensional $[RA]_i$ imaging (zebrafish embryos). Thus, absolute $[RA]_i$ values in the two-dimensional image can be obtained from the ratios by using the curves shown in Fig. 1b. RA, retinol, retinal and DEAB were administered with 0.1% dimethylsulphoxide (DMSO). The morphology of an embryo was observed with the confocal microscope simultaneously with $[RA]_i$ or with a stereomicroscope afterwards. The transgenic zebrafish line (Tg(RARE-gata2:NTD-eYFP)Id1)²¹ was obtained from the Zebrafish International Resource Center (University of Oregon). The eYFP fluorescence was observed with an inverted confocal microscope equipped with a 473-nm or 488-nm laser. cDNA encoding *raldh2*, *cyp26a1*, *cyp26b1*, *cyp26c1*, *krox20*, *hoxd4*, *otx2* or *myoD* in pCS2 was used to construct the probe for *in situ* hybridization. All embryos were allowed to grow after live $[RA]_i$ imaging, unless they were subjected to *in situ* hybridization. Embryos developed without any morphological aberrations, indicating that GEPRAs expression and light exposure are not morphogenic.

Full Methods and any associated references are available in the online version of the paper.

Received 22 November 2012; accepted 25 February 2013.

Published online 7 April 2013.

1. Stern, C. D. & Foley, A. C. Molecular dissection of Hox gene induction and maintenance in the hindbrain. *Cell* **94**, 143–145 (1998).
2. Begemann, G., Marx, M., Mebus, K., Meyer, A. & Bastmeyer, M. Beyond the neckless phenotype: influence of reduced retinoic acid signaling on motor neuron development in the zebrafish hindbrain. *Dev. Biol.* **271**, 119–129 (2004).
3. Maves, L. & Kimmel, C. B. Dynamic and sequential patterning of the zebrafish posterior hindbrain by retinoic acid. *Dev. Biol.* **285**, 593–605 (2005).
4. Sirbu, I. O., Gresh, L., Barra, J. & Duester, G. Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development* **132**, 2611–2622 (2005).
5. Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L. & Moens, C. B. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* **134**, 177–187 (2007).
6. White, R. J., Nie, Q., Lander, A. D. & Schilling, T. F. Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. *PLoS Biol.* **5**, e304 (2007).
7. Diez del Corral, R. *et al.* Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79 (2003).
8. Moreno, T. A. & Kintner, C. Regulation of segmental patterning by retinoic acid signaling during *Xenopus* somitogenesis. *Dev. Cell* **6**, 205–218 (2004).
9. Aulehla, A. & Pourquie, O. Signaling gradients during paraxial mesoderm development. *Cold Spring Harb. Perspect. Biol.* **2**, a000869 (2010).
10. White, R. J. & Schilling, T. F. How degrading: Cyp26s in hindbrain development. *Dev. Dyn.* **237**, 2775–2790 (2008).
11. Kudoh, T., Wilson, S. W. & Dawid, I. B. Distinct roles for Fgf, Wnt and retinoic acid in posteriorizing the neural ectoderm. *Development* **129**, 4335–4346 (2002).
12. Teلمان, A. A. & Cohen, S. M. Dpp gradient formation in the *Drosophila* wing imaginal disc. *Cell* **103**, 971–980 (2000).
13. Entchev, E. V., Schwabedissen, A. & González-Gaitán, M. Gradient formation of the TGF- β homolog Dpp. *Cell* **103**, 981–991 (2000).
14. Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141–152 (2007).
15. Yu, S. R. *et al.* Fgf8 morphogen gradient forms by a source–sink mechanism with freely diffusing molecules. *Nature* **461**, 533–536 (2009).
16. Müller, P. *et al.* Differential diffusivity of Nodal and Lefty underlies a reaction–diffusion patterning system. *Science* **336**, 721–724 (2012).
17. Rhinn, M. & Dollé, P. Retinoic acid signalling during development. *Development* **139**, 843–858 (2012).
18. Crick, F. Diffusion in embryogenesis. *Nature* **225**, 420–423 (1970).
19. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
20. Eichele, G. & Thaller, C. Characterization of concentration gradients of a morphologically active retinoid in the chick limb bud. *J. Cell Biol.* **105**, 1917–1923 (1987).
21. Perz-Edwards, A., Hardison, N. L. & Linney, E. Retinoic acid-mediated gene expression in transgenic reporter zebrafish. *Dev. Biol.* **229**, 89–101 (2001).
22. Emoto, Y., Wada, H., Okamoto, H., Kudo, A. & Imai, Y. Retinoic acid-metabolizing enzyme Cyp26a1 is essential for determining territories of hindbrain and spinal cord in zebrafish. *Dev. Biol.* **278**, 415–427 (2005).
23. Sawada, A. *et al.* Fgf/MAPK signaling is a crucial positional cue in somite boundary formation. *Development* **128**, 4873–4880 (2001).
24. Dubrulle, J., McGrew, M. J. & Pourquie, O. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell* **106**, 219–232 (2001).
25. Draper, B. W., Morcos, P. A. & Kimmel, C. B. Inhibition of zebrafish *fgf8* pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* **30**, 154–156 (2001).
26. Brand, M. *et al.* Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. *Development* **123**, 179–190 (1996).
27. Reifers, F. *et al.* Fgf8 is mutated in zebrafish *acerebellar* (*ace*) mutants and is required for maintenance of midbrain–hindbrain boundary development and somitogenesis. *Development* **125**, 2381–2395 (1998).
28. Hamada, A. *et al.* Retinoic acid activates myogenesis *in vivo* through Fgf8 signalling. *Dev. Biol.* **289**, 127–140 (2006).
29. Napoli, J. L. Interactions of retinoid binding proteins and enzymes in retinoid metabolism. *Biochim. Biophys. Acta* **1440**, 139–162 (1999).
30. Haugland, R. P. & Johnson, I. D. in *Fluorescent and Luminescent Probes for Biological Activity* (ed. Mason, W. T.) 40–50 (Academic, 1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank Y. Wada, R. Aoki, M. Sugiyama, F. Picazo and members of the Brain Science Institute Research Resource Center for technical assistance; C. Yokoyama and A. Terashima for critical reading of the manuscript; the FANTOM Consortium for the cDNA clones; and the Zebrafish International Resource Center for the transgenic zebrafish. This work was partly supported by grants from Japan Ministry of Education, Culture, Sports, Science and Technology Grant-in-Aid for Scientific Research on Priority Areas ‘Fluorescence Live Imaging’ and ‘Cell Innovation’ and the Human Frontier Science Program.

Author Contributions S.S. and A.M. conceived and designed the study. S.S. performed all the experiments, analysed the data and designed the manuscript. T.I. supervised the experiments on somitogenesis. T.K. and S.H. generated transgenic zebrafish lines. A.M. designed and wrote the manuscript, and supervised the project.

Author Information DNA sequences of GEPRAs are deposited in the DNA Data Bank of Japan (DDBJ) under accession numbers AB787561–AB787563. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (matsushi@brain.riken.jp).

METHODS

Materials. RA, retinol, retinal and DEAB were purchased from Sigma-Aldrich. These compounds were dissolved in dimethylsulphoxide (DMSO) to prepare stock solutions. The compounds were administered in a final concentration of 0.1% DMSO.

Gene construction. cDNA encoding the LBD of RAR- β (FANTOM3 (ref. 31), 6820403N24) was amplified using primers containing 5' SphI and 3' SacI sites. The digested product was ligated to a BamHI/SphI fragment encoding CFP and a SacI/XhoI fragment encoding YFP. The ligated product was subcloned into the BamHI/XhoI sites of the pCS2 vector for mammalian expression. In this plasmid, genes encoding the donor and/or acceptor were replaced to improve the RA indicator. Similarly, cDNA encoding RAR- γ (FANTOM3 (ref. 31), F730319B19) was used to make a SphI/SacI fragment encoding the LBD. Together with the BamHI/SphI fragment encoding CFP and a SacI/EcoRI fragment encoding YFP, cDNA coding for the indicator was subcloned into the BamHI/EcoRI sites of the pCS2 vector. Introduction of two amino-acid substitutions (R269A and S280A)³² in the LBD of RAR- β was performed as described previously³³.

Characterization of the indicators. HeLa cells were grown in DMEM medium supplemented with 10% FBS. cDNAs were transfected into cells by using Lipofectamine and Plus reagent (Invitrogen). Two days after transfection, the medium was exchanged with DMEM/F12 without phenol red. Dose-response curves were determined with nine transfected cells.

Generation of GEPRAs transgenic zebrafish lines. GEPRAs cDNAs were subcloned into the pT2KXIGAIin vector, and transgenic lines were created as described previously³⁴. Fish were maintained at 28.0 °C.

In vivo RA imaging. A glass bead (Iuchi BZ-1) was placed on a coverslip, and 1% agarose (Takara L03) in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.4 mM CaCl₂, 0.16 mM MgSO₄) was poured on the coverslip and allowed to harden. The glass bead was then removed to generate a round chamber. An embryo that had been anaesthetized with Tricaine at more than 16 hpf was placed in the chamber and covered with 0.3% agarose in E3 medium. The chamber was submerged in E3 medium containing Tricaine. Time-lapse two-dimensional imaging was performed in the *xy-t* mode with an FV1000 (Olympus) confocal inverted microscope system equipped with a $\times 10$ objective lens (numerical aperture 0.4; UPlanApo) and a 440-nm laser. With the aperture fully open, non-confocal fluorescence images were acquired. The ratio imaging of embryos was performed under exactly the same conditions as for calibration experiments with HeLa cells. Thus, absolute [RA]_i values in the image can be obtained from the ratios using the curves shown in Fig. 1b. Image processing was performed with ImageJ software (<http://rsb.info.nih.gov/ij/>). After the background had been subtracted, ratio images were generated and median-filtered.

Estimation of GEPRAs concentration in embryos. Confocal fluorescence images were acquired with a GEPRAs transgenic embryo and a series of YFP solutions of various concentrations with excitation at 488 nm. The size of the confocal aperture was about 2 Airy disks. By comparing the intensities of the confocal fluorescence images, the GEPRAs concentration in the embryo was estimated to be approximately 0.1 μ M. Considering the concentration³⁵ and affinity for RA (refs 36, 37) of

cellular RA-binding proteins, which were reported to be 10 μ M and 0.13–2 nM, respectively, only a small fraction of RA molecules should be bound to GEPRAs.

Transgenic RARE-eYFP zebrafish. Transgenic zebrafish (RARE-gata2:NTD-eYFP)l₁²¹ were obtained from the Zebrafish International Resource Center (University of Oregon). Fluorescence was observed with an FV1000 confocal microscope (Olympus) with a 488-nm laser or an FV10i confocal microscope (Olympus) with a 473-nm laser.

Depletion of endogenous RA and application of exogenous RA. Embryos were incubated in E3 medium containing 0.1% DMSO (vehicle), 10 μ M DEAB, 1 nM RA with 10 μ M DEAB, or 10 nM RA with 10 μ M DEAB from 4 to 24 hpf. After 24 hpf, embryos were bathed in E3 medium. RA imaging was performed at 11 hpf (roughly the three-somite stage). The imaged embryos were kept at 28.0 °C until their morphologies were examined at 36 hpf with a stereomicroscope (MZ16F; Leica) equipped with a charge-coupled device (CCD) camera (DP50; Olympus).

In situ hybridization. cDNA encoding *raldh2*, *cyp26a1*, *cyp26b1*, *cyp26c1*, *krox20*, *hoxd4*, *otx2* or *myoD* was amplified from a cDNA library prepared from 10-hpf or 24-hpf embryos and cloned into the pCS2 vector. Probes were labelled with a digoxigenin or fluorescein labelling mix (Roche Diagnostics) and detected with alkaline phosphatase-conjugated antibodies (1:5,000; Roche Diagnostics). The colorimetric reaction was performed with BM purple reagent (Roche Diagnostics). Images were captured with a stereomicroscope (MZ16F; Leica) equipped with a CCD camera (DP50; Olympus). Fluorescence and *in situ* hybridization images were linearly registered with the Image Processing Toolbox in MATLAB software (MathWorks).

Knock down of *fgf8*. Splice-site-targeted morpholino oligonucleotides for the *fgf8* gene (E2I2 (5'-TAGGATGCTCTTACCATGAACGTCG-3') and E3I3 (5'-CAC ATACCTTGCCAATCAGTTTCCC-3')) were used to block its pre-mRNA splicing in zebrafish embryos²⁵. One-cell-stage embryos were injected with both E2I2 and E3I3 (2.5 ng of each) or a control oligonucleotide (5 ng; 5'-CCTCTTACCTCAGTTACAATTATA-3').

- Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
- Zhang, Z. P. *et al.* Role of Ser289 in RAR γ and its homologous amino acid residue of RAR α and RAR β in the binding of retinoic acid. *Arch. Biochem. Biophys.* **380**, 339–346 (2000).
- Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **15**, e78 (2000).
- Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the Tol2 transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genesis* **174**, 639–649 (2006).
- Gustafson, A. L., Donovan, M., Annerwall, E., Dencker, L. & Eriksson, U. Nuclear import of cellular retinoic acid-binding protein type I in mouse embryonic cells. *Mech. Dev.* **58**, 27–38 (1996).
- Dong, D., Ruuska, S. E., Levinthal, D. J. & Noy, N. Distinct roles for cellular retinoic acid-binding proteins I and II in regulating signaling by retinoic acid. *J. Biol. Chem.* **274**, 23695–23698 (1999).
- Norris, A. W., Cheng, L., Giguère, V., Rosenberger, M. & Li, E. Measurement of subnanomolar retinoic acid binding affinities for cellular retinoic acid binding proteins by fluorometric titration. *Biochim. Biophys. Acta* **1209**, 10–18 (1994).

A pathogenic picornavirus acquires an envelope by hijacking cellular membranes

Zongdi Feng¹, Lucinda Hensley¹, Kevin L. McKnight¹, Fengyu Hu¹, Victoria Madden², LiFang Ping¹, Sook-Hyang Jeong³, Christopher Walker⁴, Robert E. Lanford⁵ & Stanley M. Lemon^{1,6,7}

Animal viruses are broadly categorized structurally by the presence or absence of an envelope composed of a lipid bilayer membrane¹, attributes that profoundly affect stability, transmission and immune recognition. Among those lacking an envelope, the *Picornaviridae* are a large and diverse family of positive-strand RNA viruses that includes hepatitis A virus (HAV), an ancient human pathogen that remains a common cause of enterically transmitted hepatitis^{2–4}. HAV infects in a stealth-like manner and replicates efficiently in the liver⁵. Virus-specific antibodies appear only after 3–4 weeks of infection, and typically herald its resolution^{3,4}. Although unexplained mechanistically, both anti-HAV antibody and inactivated whole-virus vaccines prevent disease when administered as late as 2 weeks after exposure⁶, when virus replication is well established in the liver⁵. Here we show that HAV released from cells is cloaked in host-derived membranes, thereby protecting the virion from antibody-mediated neutralization. These enveloped viruses ('eHAV') resemble exosomes⁷, small vesicles that are increasingly recognized to be important in intercellular communications. They are fully infectious, sensitive to extraction with chloroform, and circulate in the blood of infected humans. Their biogenesis is dependent on host proteins associated with endosomal-sorting complexes required for transport (ESCRT)⁸, namely VPS4B and ALIX. Whereas the hijacking of membranes by HAV facilitates escape from neutralizing antibodies and probably promotes virus spread within the liver, anti-capsid antibodies restrict replication after infection with eHAV, suggesting a possible explanation for prophylaxis after exposure. Membrane hijacking by HAV blurs the classic distinction between 'enveloped' and 'non-enveloped' viruses and has broad implications for mechanisms of viral egress from infected cells as well as host immune responses.

Supernatant fluids of hepatoma cell cultures infected with low-passage, non-cytopathic HAV⁹ (Fig. 1A) contain two populations of virus particles that are resolved in isopycnic iodixanol gradients (Fig. 1B). One bands at a low density consistent with membrane association ($1.06\text{--}1.10\text{ g cm}^{-3}$, fractions 8–12) and is not detected in a capsid antigen enzyme-linked immunosorbent assay (ELISA) (Fig. 1C, left), whereas the other bands at the density expected for picornaviruses ($1.22\text{--}1.28\text{ g cm}^{-3}$, fractions 18–22) and is readily detected by ELISA. Electron microscopy of the light fractions revealed numerous virus-like particles enclosed in membranes (see a–d in Fig. 1D, and Supplementary Fig. 1a) with morphology indistinguishable from $\sim 27\text{-nm}$ HAV particles in dense fractions (see e in Fig. 1D). These membrane-bound structures ranged from 50 to 110 nm in diameter, a size similar to that of exosomes⁷, and contained one to four virus-like particles (Supplementary Fig. 1b). Consistent with this, viral RNA banded in gradients with capsid protein (VP2) and the exosome-associated protein flotillin-1 (Supplementary Fig. 1c).

A modified plaque assay (infrared fluorescence immunofocus assay; IR-FIFA)¹⁰ revealed the membrane-wrapped particles to be infectious (Fig. 1E) with a specific infectivity equivalent to that of virions (Fig. 1F). Extraction with chloroform, a classic method for distinguishing enveloped from non-enveloped viruses, had no effect on standard virions but resulted in a $2\log_{10}$ decrease in infectious virus in the light fraction (Fig. 1E and Supplementary Fig. 2a). We call these membrane-wrapped HAV particles 'enveloped HAV' (eHAV). Capsid antigen could be detected in the eHAV fraction after treatment with a nonionic detergent (1% Igepal CA-630; NP-40) (Fig. 1C, right). This shifted the viral particles to an intermediate density in iodixanol gradients ($1.15\text{--}1.17\text{ g cm}^{-3}$) but did not destroy infectivity (Supplementary Fig. 2b, c). A potent, neutralizing, monoclonal antibody (mAb), K24F2 (ref. 11), failed to neutralize eHAV (Fig. 1G), providing further evidence for complete envelopment of the capsid.

In 12 experiments, eHAV represented $79 \pm 13\%$ (mean \pm s.d.) of virus in medium from infected cell cultures. eHAV was released from multiple cell types and was also observed with high-passage cytopathic virus (Supplementary Fig. 3a–e). Gradient-purified eHAV contained mostly mature VP2 (Fig. 1H, lane 3, and Supplementary Fig. 2d), indicating that enveloped virions had undergone maturation cleavage of VP0 to VP4 + VP2. However, whereas non-enveloped virions contained fully processed VP1, eHAV contained primarily unprocessed VP1pX (Fig. 1H; compare lanes 3 and 4). pX is an unusual 8-kDa carboxy-terminal extension on VP1 that is unique to HAV among picornaviruses (Fig. 1A). It functions in virion assembly and is processed from VP1 by an unidentified host protease late in the viral lifecycle^{12,13}. pX was protected from proteinase K in eHAV particles, but was rendered susceptible to digestion after treatment with NP-40 (Supplementary Fig. 2e). Thus, pX is fully enclosed in membranes. We conclude that most HAV is released enveloped in host membranes, a process we term 'membrane hijacking'.

Infectious virus circulating in the blood of infected humans and chimpanzees (*Pan troglodytes*) possesses the buoyant density of eHAV (Fig. 2a and Supplementary Fig. 4a–e), confirming the relevance of these observations to HAV pathogenesis. However, virus in faeces is non-enveloped (Fig. 2b and Supplementary Fig. 4c). Because most faecal virus is produced in the liver^{14,15}, the membrane may be stripped from eHAV during passage through the biliary tract to the gut. However, suspending eHAV in bile minimally altered its buoyant density (Supplementary Fig. 5a).

It has been suggested¹⁶ that poliovirus, a distantly related picornavirus, could be released from cells wrapped in a single membrane if entrapped in an autophagosome that fuses with the plasma membrane. However, RNA-mediated interference knockdown of beclin-1, which mediates autophagosome formation, did not inhibit eHAV release (Supplementary Fig. 6a, b). In contrast, knockdown of VPS4B or

¹Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7292, USA. ²Department of Pathology and Laboratory Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7292, USA. ³Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam-si, Gyeonggi-do, 463-707, South Korea. ⁴Center for Vaccines and Immunity, The Research Institute at The Nationwide Children's Hospital, and Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, Ohio 43205, USA. ⁵Department of Virology and Immunology, Texas Biomedical Research Institute, San Antonio, Texas 78227, USA. ⁶Division of Infectious Disease, Department of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7292, USA. ⁷Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7292, USA.

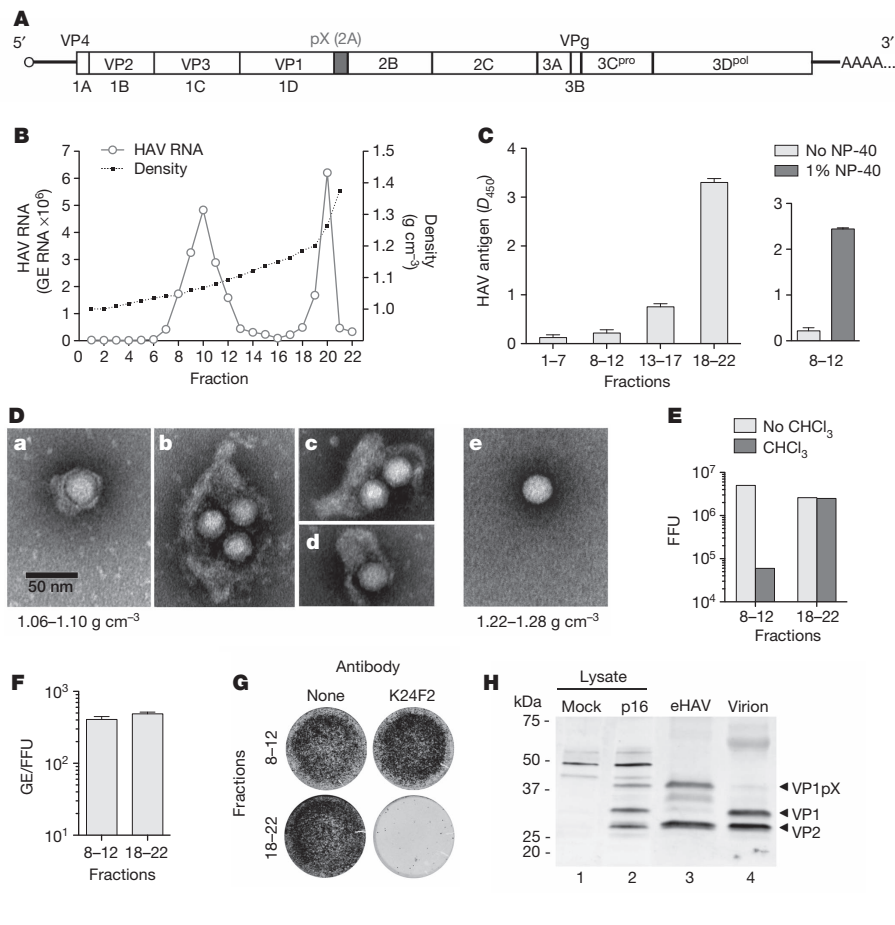


Figure 1 | Enveloped particles (eHAV) are the dominant form of virus released from infected cell cultures. **A**, HAV genome organization. The polyprotein is depicted as a box with pX highlighted. **B**, Buoyant density of HAV particles released by Huh-7.5 cells in iodixanol gradients. eHAV bands at $1.06\text{--}1.10\text{ g cm}^{-3}$, whereas non-enveloped HAV bands at $1.22\text{--}1.28\text{ g cm}^{-3}$. GE, genome equivalents. **C**, Left, HAV capsid antigen was detected by ELISA only in pools of denser fractions from the gradient in **B**. Right, capsid antigen was detected in light fractions after treatment with 1% NP-40. Data shown are D_{450} (mean and range) in duplicate assays. **D**, Electron microscopic images of negatively stained eHAV (**a–d** from fraction 10 in **B**) and a non-enveloped virion (**e**, fraction 20 in **B**). **E**, Infectious titre of pooled fractions containing eHAV or non-enveloped virions before and after extraction with chloroform. FFU, focus-forming units. **F**, Specific infectivity of pooled fractions containing eHAV or non-enveloped virions, calculated by dividing the HAV RNA copy number (GE, qRT-PCR) by the infectious titre (FFU, IR-FIFA). Values shown are means and range for duplicate RT-PCR reactions. **G**, eHAV is resistant to neutralization by anti-capsid monoclonal antibody K24F2 (ref. 11). Antibody–virus mixtures were incubated for 1 h at 37°C and inoculated onto cells for 1 h, followed by removal of the inoculum, washing three times with PBS, and the addition of an agarose overlay. Viral antigen was detected by IR-FIFA¹⁰. **H**, Immunoblots of HAV capsid proteins (VP1 and VP2) in lysates of mock-infected or HAV-infected cells (lanes 1 and 2), gradient-purified eHAV (lane 3) and chloroform-extracted non-enveloped virions (lane 4).

ALIX, which are ESCRT-III-associated proteins that facilitate the budding of many enveloped viruses^{17–20}, inhibited the release of eHAV from cells (Fig. 3a–c and Supplementary Figs 6c and 7a) but not viral RNA replication (Fig. 3d), intracellular assembly of capsids or encapsidation of viral RNA (Supplementary Fig. 7a–c). ESCRT complexes act sequentially to sort and load cargo into multivesicular bodies (MVBs)⁸. VPS4B is an ATPase that provides energy for the dissolution

of ESCRT-III complexes at membrane fission, whereas ALIX contributes to exosome biogenesis and also facilitates lentivirus budding^{8,19,21}. Knockdown of VPS4B and ALIX inhibited the release of both enveloped and non-enveloped HAV (Fig. 3c), suggesting these share a common egress pathway and that non-enveloped virus in extracellular fluids might be derived from eHAV. Knockdown of HRS and TSG101, which are ESCRT-0 and ESCRT-I proteins involved early in cargo recruitment, did not inhibit eHAV release (Fig. 3b, c). Thus, eHAV release is dependent on ESCRT-associated proteins but not on the entire ESCRT machinery, in a similar manner to the budding of some enveloped viruses^{8,18}. Although an interaction of ALIX with the ESCRT-III protein CHMP4B facilitates the budding of human immunodeficiency virus²², CHMP4A, CHMP4B or CHMP4C knockdown minimally decreased eHAV release (Supplementary Fig. 8).

Structural proteins of enveloped viruses interact with ESCRT-related proteins by means of Pro-rich late (‘L’) domain motifs: PPXY, P(S/T)AP or (L)YPX_{1/3}L (refs 8, 17). Only two domains in the 2,227-residue-long HAV polyprotein conform to these motifs: conserved, tandem YPX₃L motifs separated by 28 residues in the VP2 capsid protein (Fig. 3e). YPX_{1/3}L motifs mediate interactions with ALIX (ref. 19), and their presence in VP2 is consistent with a requirement for ALIX. Disrupting either

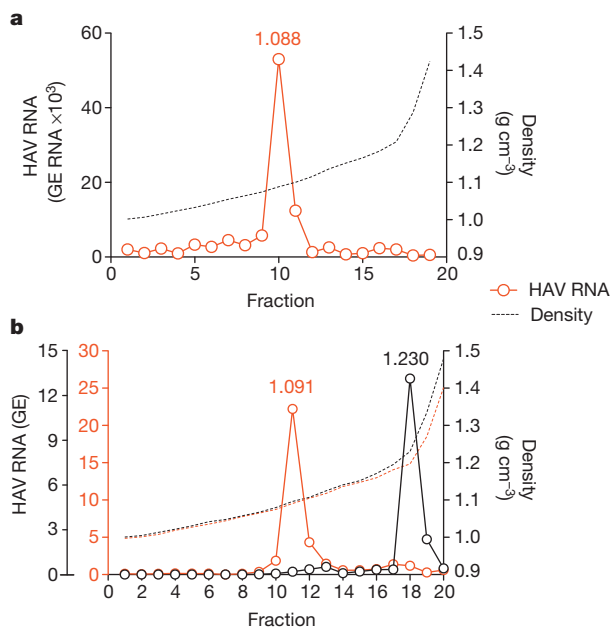


Figure 2 | eHAV circulates in the blood of HAV-infected humans and chimpanzees. **a**, Distribution of HAV RNA in an iodixanol gradient loaded with early, acute-phase serum from patient BH12. **b**, Buoyant density of HAV particles from plasma (red open circles) and faeces (black open circles) of an experimentally infected chimpanzee, x0293 (ref. 5). Faecal HAV RNA is shown $\times 10^5$; plasma RNA is shown $\times 10^2$ (also see Supplementary Fig. 4). The low buoyant density of circulating virus was not due to a passive effect of serum, because the buoyant density of non-enveloped virions was not altered by suspension in 90% serum (data not shown).

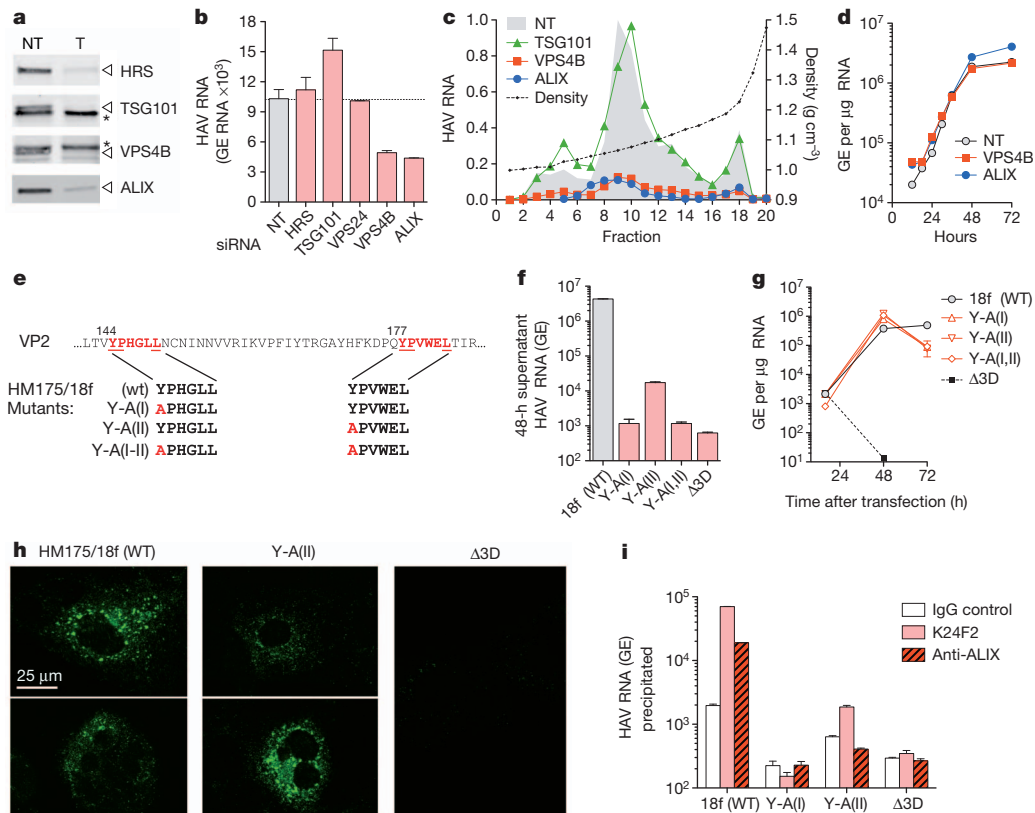


Figure 3 | eHAV biogenesis requires ESCRT-associated proteins.

a, Immunoblots of HRS, TSG101, VPS4B and ALIX in HAV-infected cells 72 h after transfection with the indicated gene-specific targeting (T) or non-targeting (NT) control siRNAs. Asterisks, non-specific bands. **b**, qRT-PCR assays confirmed knockdown of TSG101 and VPS4B (Supplementary Fig. 6c). **c**, Viral RNA detected in culture fluids 48–72 h after siRNA transfection. **d**, Relative yield and buoyant density of virus released from siRNA transfected cells (as in **a**) determined in iodixanol gradients. RNA associated with eHAV particles (fractions 8–12) was decreased by 85% and 87% by depletion of VPS4B and ALIX, respectively. **e**, Huh-7.5 cells were transfected with the indicated siRNAs for 3 days, then infected with gradient-purified eHAV at a multiplicity of infection of 20. Cell-associated HAV RNA was assayed by qRT-PCR. Knockdown efficiency was confirmed at the end of the experiment (not shown). **f**, Tandem YPX₃L ALIX-interacting motifs (red)

Y-PX₃L motif by substituting Ala for Tyr (Fig. 3e) severely decreased virus release but did not impair viral RNA replication (Fig. 3f, g). Intracellular capsid antigen was similarly decreased, suggesting that the mutations impede capsid assembly, but the antigen was detected in cells transfected with Y-A(II) RNA in which the second YPX₃L motif had been ablated (Fig. 3h). Limited capsid assembly was confirmed by K24F2 precipitation of RNase-protected viral RNA from detergent-treated cell lysates (Fig. 3i). Consistent with a role for ALIX in eHAV biogenesis, anti-ALIX antibody precipitated encapsidated (RNase-protected) viral RNA from lysates of cells transfected with wild-type, but not Y-A(II), RNA (Fig. 3i). Furthermore, ALIX banded at the same density as eHAV in iodixanol gradients (Supplementary Fig. 7d). Although details remain to be worked out, these data suggest that HAV co-opts ALIX to facilitate its envelopment in host membranes. Where this occurs is uncertain, but HAV particles have been observed within cytoplasmic vesicles in infected liver^{14,23}.

The cloaking of its capsid by membranes allows eHAV to evade neutralizing antibodies (Fig. 1G and Supplementary Fig. 9a) and explains how infectious virus co-exists with anti-HAV in serum¹⁵. However, antibodies protect against hepatitis A when passively transferred after HAV replication has been well established in the liver^{5,6}. We therefore reasoned that eHAV might be neutralized after binding to or entering hepatocytes. Consistent with this, eHAV was neutralized

in VP2. Below are the motif sequences in wild-type (WT, HM175/18f) and mutant viral RNAs. **f**, Viral RNA released into culture fluids 24–48 h after electroporation of Huh-7.5 cells with wild-type and mutant viral RNAs. Δ 3D is a replication-incompetent subgenomic replicon RNA with a frameshift mutation in 3D^{pol}. **g**, Intracellular HAV RNA after electroporation of cells with WT and VP2 mutant RNAs. **h**, Confocal microscopy showing K24F2 detection of capsid antigen in cells electroporated 48 h previously with WT or mutant RNAs. K24F2 recognizes a conformation-dependent assembled neutralization epitope in the viral capsid¹¹. **i**, Anti-capsid (K24F2) and anti-ALIX (Bethyl) antibody-mediated immunoprecipitation of encapsidated, RNase-resistant viral RNA in detergent-treated lysates of cells electroporated with wild-type or mutant HAV RNAs. RT-PCR data are means \pm s.e.m. for two or three replicate assays; all results are representative of two or three independent experiments.

when cells were exposed to antibody either immediately before inoculation with virus or up to 6 h afterwards (Fig. 4a, b). IgG and IgA anti-capsid mAbs neutralized eHAV when added to cultures after removal of the inoculum, but an IgM mAb, H14C42 (ref. 11), failed to do so despite potent neutralizing activity against non-enveloped HAV (Supplementary Fig. 9b). Taken together, these results show capsid is the target for neutralization and that eHAV membranes must rupture before neutralization, but not before eHAV is internalized and inaccessible to H14C42 in the medium. This suggests that eHAV may be neutralized after endocytosis. Chloroquine efficiently blocked infection by eHAV (but not by non-enveloped HAV) (Fig. 4c), indicating that endosomal acidification is important for eHAV entry, whereas an antibody against the HAV receptor, TIM-1 (HAVCR-1)²⁴, inhibited both (Fig. 4d). We found eHAV membranes to be stable at pH 5.0 (Supplementary Fig. 5b), but they could nonetheless be degraded in late endosomes and lysosomes²⁵. If so, this would render the capsid accessible to TIM-1, which undergoes constitutive endocytosis and trafficking to late endosomes and lysosomes²⁶, as well as to anti-HAV. Many details of eHAV entry and neutralization remain to be explained. However, TRIM-21 (tripartite motif-containing 21), which is implicated in intracellular neutralization of other viruses²⁷, seems to have no function in neutralization of eHAV (Supplementary Fig. 9c).

Our findings show how a 'non-enveloped' picornavirus promotes its egress in the absence of cell lysis, and reveal a previously unrecognized strategy by which a virus cloaks itself in host membranes to evade neutralizing antibodies. The host membrane enveloping eHAV is

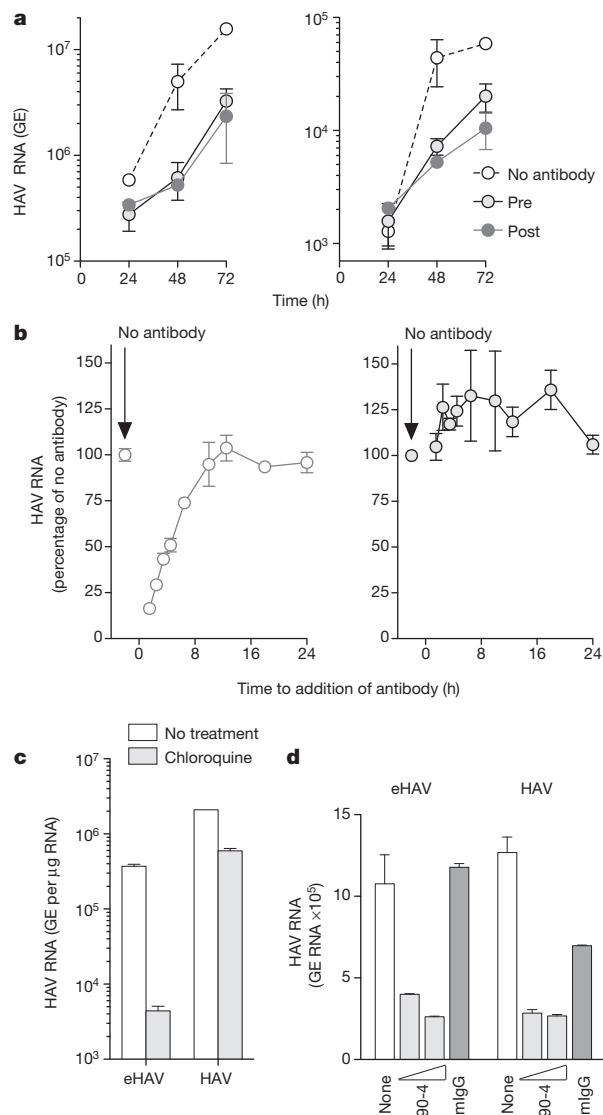


Figure 4 | Extracellular eHAV is resistant to antibody-mediated neutralization but is neutralized by antibody after infection.

a, Neutralization of eHAV in cells treated with anti-HAV before (pre) or after (post) adsorption of virus. Left, intracellular; right, extracellular. Pretreated cells were incubated with JC plasma (1:50–1:100) for 1 h, washed extensively with PBS, then incubated with gradient-purified eHAV (multiplicity of infection about 6) for 1 h in the absence of antibody. For post-treatment, JC was added to the medium on refeeding after removal of the inoculum. Cells were subsequently re-fed at 24 h intervals with fresh medium containing no antibody. Data are means \pm s.e.m. from duplicate cultures and are representative of six independent experiments. **b**, Adding anti-HAV to medium as late as 6 h after removal of the inoculum restricts the replication of eHAV (left) but not that of non-enveloped HAV (right). JC antibody was added at intervals after removal of the inoculum, and intracellular viral RNA was quantified at 48 h. Results are means \pm s.e.m. of replicate RT-PCR assays and are representative of three independent experiments. **c**, Chloroquine (50 μ M) blocks eHAV entry but not HAV entry. Cells were treated for 30 min before inoculation with virus, and were harvested at 48 h to assay intracellular HAV RNA. **d**, Anti-TIM-1 mAb 190-4 blocks entry of both eHAV and HAV. GL37 cells were incubated with 190-4 (10–100 μ g ml $^{-1}$) or control IgG (100 μ g ml $^{-1}$) for 1 h at 37 $^{\circ}$ C, then inoculated with virus for 1 h. Intracellular HAV RNA was assayed at 72 h. RT-PCR data are means \pm s.e.m. for two or three replicate assays; all results are representative of two or three independent experiments.

likely to facilitate its spread within the liver. Antibodies restrict the replication of eHAV when added to cultures several hours after infection, but further studies will be needed to determine how this happens and whether this can explain why immune globulin and vaccines protect against hepatitis when administered long after exposure, when virus is already circulating in blood⁵. Our results suggest that categorizing viruses into those that are enveloped and those that are not, long a tradition in virology¹, is overly simplistic, and that efforts should be made to identify other examples of membrane hijacking by 'non-enveloped' viruses.

METHODS SUMMARY

Non-cytopathic, cell-culture-adapted HM175/p16 HAV⁹ was propagated in Huh-7.5 cells. Reverse molecular genetics studies were performed with pHM175/18f, a molecular clone of a related cytopathic variant²⁸. Buoyant density was assessed in 8–40% iodixanol (Opti-Prep) gradients centrifuged at 141,000g for 48 h at 4 $^{\circ}$ C. Viral RNA was measured by quantitative PCR with reverse transcription (qRT-PCR) with primers targeting the 5' untranslated region. Infectivity was quantified by IR-FIFA¹⁰. For RNA-mediated interference studies, cells were transfected with SmartPool short interfering RNAs (siRNAs) (Dharmacon) and samples were collected 48–72 h later for viral RNA quantification. To analyse VP2–ALIX interactions, cell lysates were prepared 48 h after electroporation of mutant and wild-type viral RNAs, treated with RNase and immunoprecipitated. RNA extracted from immunoprecipitates was assayed by HAV-specific qRT-PCR. For intracellular neutralization, cells were incubated with eHAV for 1 h at 37 $^{\circ}$ C, then washed extensively. Antibodies were added at intervals, and intracellular and extracellular HAV RNA was quantified at 48–72 h. For standard neutralization assays, virus was incubated with antibodies for 1 h at 37 $^{\circ}$ C, then inoculated onto cells for IR-FIFA¹⁰.

Full Methods and any associated references are available in the online version of the paper.

Received 9 August 2012; accepted 20 February 2013.

Published online 31 March 2013.

- Harrison, S. C. in *Fields Virology* (eds Knipe, D. M. et al.) Ch. 3 59–98 (Lippincott Williams & Wilkins, 2007).
- Feng, Z. & Lemon, S. M. in *The Picornaviruses* (eds Ehrenfeld, E., Domingo, E. & Roos, R. P.) Ch. 25 383–396 (ASM Press, 2010).
- Lemon, S. M. Type A viral hepatitis: new developments in an old disease. *N. Engl. J. Med.* **313**, 1059–1067 (1985).
- Martin, A. & Lemon, S. M. in *Hepatitis Viruses* (ed. Ou, J.) 23–50 (Kluwer Academic, 2002).
- Lanford, R. E. et al. Acute hepatitis A virus infection is associated with a limited type I interferon response and persistence of intrahepatic viral RNA. *Proc. Natl Acad. Sci. USA* **108**, 11223–11228 (2011).
- Victor, J. C. et al. Hepatitis A vaccine versus immune globulin for postexposure prophylaxis. *N. Engl. J. Med.* **357**, 1685–1694 (2007).
- Bobrie, A., Colombo, M., Raposo, G. & Thery, C. Exosome secretion: molecular mechanisms and roles in immune responses. *Traffic* **12**, 1659–1668 (2011).
- Hurley, J. H. The ESCRT complexes. *Crit. Rev. Biochem. Mol. Biol.* **45**, 463–487 (2010).
- Jansen, R. W., Newbold, J. E. & Lemon, S. M. Complete nucleotide sequence of a cell culture-adapted variant of hepatitis A virus: comparison with wild-type virus with restricted capacity for *in vitro* replication. *Virology* **163**, 299–307 (1988).
- Counihan, N. A., Daniel, L. M., Chojnacki, J. & Anderson, D. A. Infrared fluorescent immunofocus assay (IR-FIFA) for the quantitation of non-cytopathic and minimally cytopathic viruses. *J. Virol. Methods* **133**, 62–69 (2006).
- Ping, L. H. & Lemon, S. M. Antigenic structure of human hepatitis A virus defined by analysis of escape mutants selected against murine monoclonal antibodies. *J. Virol.* **66**, 2208–2216 (1992).
- Cohen, L., Benichou, D. & Martin, A. Analysis of deletion mutants indicates that the 2A polypeptide of hepatitis A virus participates in virion morphogenesis. *J. Virol.* **76**, 7495–7505 (2002).
- Graff, J. et al. Hepatitis A virus capsid protein VP1 has a heterogeneous C terminus. *J. Virol.* **73**, 6015–6023 (1999).
- Schulman, A. N. et al. Hepatitis A antigen particles in liver, bile, and stool of chimpanzees. *J. Infect. Dis.* **134**, 80–84 (1976).
- Asher, L. V. S. et al. Pathogenesis of hepatitis A in orally inoculated owl monkeys (*Aotus trivirgatus*). *J. Med. Virol.* **47**, 260–268 (1995).
- Taylor, M. P., Burgon, T. B., Kirkegaard, K. & Jackson, W. T. Role of microtubules in extracellular release of poliovirus. *J. Virol.* **83**, 6599–6609 (2009).
- Ren, X. & Hurley, J. H. Proline-rich regions and motifs in trafficking: from ESCRT interaction to viral exploitation. *Traffic* **12**, 1282–1290 (2011).
- Chen, B. J. & Lamb, R. A. Mechanisms for enveloped virus budding: can some viruses do without an ESCRT? *Virology* **372**, 221–232 (2008).
- Sette, P. et al. The Phe105 loop of Alix Bro1 domain plays a key role in HIV-1 release. *Structure* **19**, 1485–1495 (2011).
- Fujii, K., Hurley, J. H. & Freed, E. O. Beyond Tsg101: the role of Alix in 'ESCRTing' HIV-1. *Nature Rev. Microbiol.* **5**, 912–916 (2007).

21. Baietti, M. F. *et al.* Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nature Cell Biol.* **14**, 677–685 (2012).
22. Morita, E. *et al.* ESCRT-III protein requirements for HIV-1 budding. *Cell Host Microbe* **9**, 235–242 (2011).
23. Shimizu, Y. K. *et al.* Detection of hepatitis A antigen in human liver. *Infect. Immun.* **36**, 320–324 (1982).
24. Feigelstock, D. *et al.* The human homolog of HAVcr-1 codes for a hepatitis A virus cellular receptor. *J. Virol.* **72**, 6621–6628 (1998).
25. Kolter, T. & Sandhoff, K. Lysosomal degradation of membrane lipids. *FEBS Lett.* **584**, 1700–1712 (2010).
26. Balasubramanian, S. *et al.* TIM family proteins promote the lysosomal degradation of the nuclear receptor NUR77. *Sci. Signal.* **5**, ra90 (2012).
27. Mallery, D. L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc. Natl Acad. Sci. USA* **107**, 19985–19990 (2010).
28. Zhang, H. C. *et al.* An infectious cDNA clone of a cytopathic hepatitis A virus: genomic regions associated with rapid replication and cytopathic effect. *Virology* **212**, 686–697 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Duncan for helpful discussions; staff of the Michael Hooker Microscopy Facility for assistance with confocal microscopy; G. Kaplan for the gift of 190-4 antibody and GL37 cells; M. Fried for anti-HAV-negative human sera; A. Blikslager for porcine bile; and C. Rice for Huh-7.5 cells. This work was supported in part by a grant from the National Institutes of Health (R01-AI103083) and the University Cancer Research Fund.

Author Contributions Z.F., C.H., K.L.M., F.H., V.M. and L.-F.P. designed experiments, collected and analysed data. S.-H.J., C.W. and R.L. provided unique human and primate samples for analysis. Z.F. and S.M.L. conceived the study, analysed data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.L. (smlemon@med.unc.edu).

METHODS

Reagents and antibodies. Chemical reagents were purchased from Sigma unless otherwise noted. A human convalescent plasma (JC plasma) sample with high-titre anti-HAV antibody, collected 90 days after the onset of acute hepatitis A, was described previously^{11,29}. Other antibodies were obtained as follows: anti-TSG101 (Abcam), anti-HRS (Bethyl), anti-ALIX (Pierce and Bethyl), anti-VPS4B (Sigma), anti-beclin-1 (Cell Signaling), anti-flotillin-1 (BD Biosciences) and anti-HAV mAbs K24F2 and K32F2 (Commonwealth Serum Laboratories, Victoria, Australia). The sources of other anti-HAV mAbs have been described previously¹¹. Anti-TIM-1 mAb 190-4 was a gift from G. Kaplan. Guinea-pig anti-peptide antibodies against HAV VP1 (residues 211–228) and VP2 (residues 240–259) were a gift from D. Sangar.

Cells and viruses. Huh-7.5 and HepG2 (both human hepatoma cells), MRC-5 (human lung fibroblasts), FRhK-4 (fetal rhesus monkey kidney cells), BS-C-1 and GL37 (ref. 30) (both African green monkey kidney) cells were cultured in DMEM medium supplemented with 10% FBS. Cell-culture adapted variants of the HM175 strain of HAV, HM175/p16 (refs 9, 31) (non-cytopathic) and HM175/18f (refs 28, 32) (rapidly replicating, cytopathic), have been described previously^{9,32}. Infectious virus titres were determined by a modification of the IR-FIFA described previously¹⁰ performed in FRhK-4 cells.

Human samples and clinical information. All human samples and related clinical information were obtained with the approval of the Institutional Review Boards of the Seoul National University Bundang Hospital and the University of North Carolina at Chapel Hill.

Chimpanzee samples. All chimpanzee materials were archived samples collected during previous studies⁵ conducted prior to 15 December 2011.

HAV capsid antigen ELISA assay. HAV capsid antigen was detected with an ELISA modification of a previously described radioimmunoassay³³. Immobilized antigen captured by JC antibody was detected by sequential incubation of ELISA plates with K24F2 (1:1,000 dilution) and horseradish peroxidase-conjugated goat-anti-mouse antibody (1:10,000) at room temperature (20–23 °C) for 1 h each. After washing with PBS-T and substrate (3,3',5,5'-tetramethylbenzidine) addition, D_{450} was determined with a Synergy 2 (BioTek) microplate reader.

qRT-PCR assay for HAV RNA. Total cellular RNA was extracted from cell lysates with the RNeasy Kit (Qiagen); RNA from culture supernatants and gradient fractions was extracted with the QiaAmp viral RNA Isolation Kit (Qiagen). HAV RNA was quantified by real-time qRT-PCR against a synthetic RNA standard as described previously⁵. Intracellular viral RNA was normalized to total RNA abundance or glyceraldehyde-3-phosphate dehydrogenase mRNA level determined in a SYBR green one-step qRT-PCR assay (Bio-Rad). For experiments involving electroporation of synthetic viral RNAs, a two-step qRT-PCR procedure was conducted in which complementary DNA was synthesized with (dT)₂₀ as primer, followed by RNaseH digestion and qPCR assay with primers targeting the HAV 5' untranslated RNA segment.

Virus neutralization. To assess eHAV neutralization following binding or entry of eHAV into cells, Huh-7 cells (FT3-7 subline) were incubated with gradient-purified eHAV for 1 h at 37 °C, then washed extensively and refed with fresh medium. Antibodies were added at intervals, and intracellular and extracellular HAV RNA was quantified 48–72 h after virus inoculation. For standard neutralization assays, virus was incubated with antibodies for 1 h at 37 °C, then inoculated onto FRhK-4 cells which were processed subsequently for IR-FIFA¹⁰.

Immunoblotting. Cells were lysed with CellLytic Lysis Buffer (Invitrogen) in the presence of protease inhibitor cocktail (Roche). Proteins were extracted from gradient fractions by precipitation with trichloroacetic acid. Immunoblots were performed with standard procedures and the indicated antibodies. Protein bands were detected with an Odyssey Infrared Imaging System (LI-COR Biosciences).

Immunofluorescence microscopy. Huh-7.5 cells seeded on four-well chamber slides were fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100 in PBS, and blocked with 10% goat serum. Cells were then incubated with mAb K24F2 (1:300 dilution in 1% BSA) for 1 h at room temperature, washed, and incubated with fluorescein isothiocyanate-conjugated goat anti-mouse antibody (1:300; Southern Biotech) for 1 h. Images were collected with a Zeiss 510 Meta laser-scanning confocal microscope.

Isoptic gradient centrifugation of virus. Cell-culture supernatant fluids were centrifuged at 1,000g for 10 min at 4 °C to remove cells and debris, further clarified by centrifugation twice at 10,000g for 30 min, and concentrated by ultracentrifugation at 100,000g for 1 h at 4 °C. The resulting pellet was resuspended in PBS,

loaded onto a 8–40% iodixanol (Opti-Prep) step gradient, and centrifuged at 141,000g in a Superspin 630 rotor for 48 h at 4 °C in a Sorvall Ultra-80 ultracentrifuge. Approximately 20 fractions were collected from the top of the gradient; density was determined with a refractometer.

Rate-zonal centrifugation of HAV in sucrose gradients. Lysates were prepared from HAV (HM175/p16)-infected Huh-7.5 cells, loaded onto a 7.5–45% (w/v) sucrose gradient in Hank's balanced buffer, and centrifuged at 4 °C in a Beckman SW551 rotor for 3 h at 42,000 r.p.m. (167,000g). Fractions were assayed by HAV antigen ELISA and HAV RNA-specific qRT-PCR to identify pentamers, procapsids and mature HAV virions.

Immunoprecipitation of HAV. Cells (2×10^6) were lysed in the presence of detergent with CellLytic lysis buffer (Invitrogen), treated with 50 $\mu\text{g ml}^{-1}$ RNase A (Qiagen) and 25 U ml^{-1} RNase T1 (Ambion) at 37 °C for 15 min, and incubated overnight with antigen-specific and isotype-matched antibodies at 4 °C. Protein G beads were added and incubated for a further 2 h. Beads were extensively washed with lysis buffer, and captured immunoprecipitates were subjected to RNA extraction for qRT-PCR, or immunoblotting, as above.

Virus isolation from iodixanol gradient fractions. Huh-7.5 cells in 12-well culture plates were inoculated with 10–100 μl of peak iodixanol gradient fractions containing eHAV recovered from human or chimpanzee serum or plasma. Cells were re-fed weekly, and split 1:3 at 2 weeks. RNA was extracted from supernatant culture fluids 8, 23 and 30 days after inoculation and assayed for HAV RNA by qRT-PCR as described above. To confirm viral isolates, RNA was extracted from supernatant culture fluids with the QIAamp Viral RNA Mini Kit (Qiagen), and cDNA was transcribed with random primers (Superscript III First-Strand Synthesis System; Invitrogen). HAV cDNA was amplified by nested RT-PCR with primer sets bracketing the VP1pX-2B junction (nucleotides 2908–3306 in wild-type HM-175) (PrimeSTAR; TaKaRa). Second-round PCR products were isolated from agarose gels (Mini-Elute Gel Extraction Kit; Qiagen) and directly sequenced.

Generation and characterization of HAV VP2 mutants. Mutagenesis of the infectious cDNA clone, pT7-HM175/18f, containing the complete sequence of HM175/18f virus²⁸, was accomplished with the Quick Change II Site-directed Mutagenesis Kit (Agilent). PCR-derived fragments were sequenced to ensure sequence fidelity. Plasmids were linearized and viral RNA was synthesized with the T7 Megascript Kit (Ambion). Synthetic viral RNA (10 μg) was mixed with 5×10^6 Huh-7.5 cells in a 4-mm cuvette and pulsed once at 250 V, 950 μF and 100 Ω in a Gene Pulser Xcell Total electroporation system (Bio-Rad). Cells and supernatants were harvested at subsequent intervals for RNA extraction and qRT-PCR assay.

Electron microscopy. Gradient fractions (2.5 μl each) were spread onto the surface of a glow-discharged 400 mesh Formvar/carbon-coated copper grid and allowed to adsorb for 5 min. The grid was inverted onto a droplet of 1% glutaraldehyde in 0.15 M sodium phosphate pH 7.4 for 1 min, followed by rinsing with deionized water and staining with 3% ammonium molybdate pH 7.0. The grid was air-dried before examination in a LEO EM910 transmission electron microscope (Carl Zeiss SMT, Inc.) operating at 80 kV. Digital images were acquired with a Gatan Orius SC1000 CCD Digital Camera and Digital Micrograph 3.11.0 software (Gatan, Inc.).

siRNA knockdown of protein expression. Huh-7.5 cells were transfected with gene-specific SmartPool siRNAs (Dharmacon; see Supplementary Table 1) with SilentFect (Bio-Rad) transfection reagent. Knockdown efficiency was determined at 48–72 h by qRT-PCR (SYBR Green One-step qRT-PCR Kit; BioRad) measurement of mRNAs, or by immunoblotting as described.

29. Lemon, S. M. *et al.* Specific immunoglobulin M response to hepatitis A virus determined by solid-phase radioimmunoassay. *Infect. Immun.* **28**, 927–936 (1980).
30. Feigelsstock, D., Thompson, P., Mattoo, P. & Kaplan, G. G. Polymorphisms of the hepatitis A virus cellular receptor 1 in African green monkey kidney cells result in antigenic variants that do not react with protective monoclonal antibody 190/4. *J. Virol.* **72**, 6218–6222 (1998).
31. Taylor, K. L. *et al.* Attenuation phenotype of a cell culture-adapted variant of hepatitis A virus (HM175/p16) in susceptible new world owl monkeys. *J. Infect. Dis.* **168**, 592–601 (1993).
32. Lemon, S. M. *et al.* Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination. *J. Virol.* **65**, 2056–2065 (1991).
33. Lemon, S. M. *et al.* Transmission of hepatitis A virus among recently captured Panamanian owl monkeys. *J. Med. Virol.* **10**, 25–36 (1982).

Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization

Shireen A. Sarraf¹, Malavika Raman¹, Virginia Guarani-Pereira¹, Mathew E. Sowa¹, Edward L. Huttlin¹, Steven P. Gygi¹ & J. Wade Harper¹

The PARKIN ubiquitin ligase (also known as PARK2) and its regulatory kinase PINK1 (also known as PARK6), often mutated in familial early-onset Parkinson's disease, have central roles in mitochondrial homeostasis and mitophagy^{1–3}. Whereas PARKIN is recruited to the mitochondrial outer membrane (MOM) upon depolarization via PINK1 action and can ubiquitylate porin, mitofusin and Miro proteins on the MOM^{1,4–11}, the full repertoire of PARKIN substrates—the PARKIN-dependent ubiquitylome—remains poorly defined. Here we use quantitative diGly capture proteomics (diGly)^{12,13} to elucidate the ubiquitylation site specificity and topology of PARKIN-dependent target modification in response to mitochondrial depolarization. Hundreds of dynamically regulated ubiquitylation sites in dozens of proteins were identified, with strong enrichment for MOM proteins, indicating that PARKIN dramatically alters the ubiquitylation status of the mitochondrial proteome. Using complementary interaction proteomics, we found depolarization-dependent PARKIN association with numerous MOM targets, autophagy receptors, and the proteasome. Mutation of the PARKIN active site residue C431, which has been found mutated in Parkinson's disease patients, largely disrupts these associations. Structural and topological analysis revealed extensive conservation of PARKIN-dependent ubiquitylation sites on cytoplasmic domains in vertebrate and *Drosophila melanogaster* MOM proteins. These studies provide a resource for understanding how the PINK1–PARKIN pathway re-sculpts the proteome to support mitochondrial homeostasis.

PARKIN, a RING-HECT hybrid E3 ubiquitin ligase¹⁴, is recruited to the MOM in cancer cell lines, mouse embryo fibroblasts (MEFs) and primary neurons to promote mitophagy in response to mitochondrial depolarization^{1,2,15,16}. This process requires PINK1, which binds the MOM translocase (TOMM) complex and phosphorylates PARKIN on multiple residues to promote MOM localization and ligase activation^{17–20}. The best understood PARKIN substrates are MFN1/2 and RHOT1/2 (MIRO in *Drosophila*), MOM-tethered GTPases whose PARKIN-dependent proteasomal turnover alters fission–fusion cycles and microtubule-dependent trafficking of mitochondria, respectively^{1,5–8,10,11}. The MOM porin proteins VDAC1/2/3 are also ubiquitylated by PARKIN, and are required for PARKIN localization on mitochondria through a poorly understood mechanism^{9,21}. Proteomic studies of purified mitochondria have also identified additional proteins whose abundance is either decreased or increased upon PARKIN activation or mitochondrial depolarization^{1,4}, but precisely how these proteins are regulated and the extent to which ubiquitin is involved is unclear. Consequently, we do not have a comprehensive understanding of cellular PARKIN targets, which is critical for elucidating how PARKIN promotes mitochondrial homeostasis. Moreover, for the vast majority of E3 ubiquitin ligases, including PARKIN, the extent to which ubiquitin transfer is site specific and signal dependent is largely unknown, and a global understanding of site specificity across a wide range of substrates is lacking for any E3. Such information is necessary for

decoding the cellular and molecular topology of E3 function and for defining how the ubiquitin system re-sculpts the proteome.

We set out to systematically identify cellular PARKIN-dependent ubiquitylation targets and the dynamics of modification in a site-specific manner using quantitative diGly (QdiGly) proteomics. QdiGly merges antibody-based capture of 'diGly remnant'-containing peptides (marking ubiquitylated proteins after trypsinolysis) and stable isotopic labelling with amino acids in culture (SILAC) to identify ubiquitylation sites that are dynamically induced, in this case, upon mitochondrial depolarization^{12,13}. To overcome inherent stochastic sampling for low-abundance diGly peptides (see Supplementary Fig. 1a, b)^{12,13}, we performed 73 control and QdiGly profiling experiments in four cell lines with and without elevated PARKIN levels (~tenfold) (Supplementary Fig. 2a) in a three-tiered manner, varying the length of depolarization with carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) (Fig. 1a and Supplementary Table 1). In some cases, protein turnover was blocked by proteasome inhibition with bortezomib (Btz) or autophagy inhibition with bafilomycin A (BafA) (Supplementary Table 1). Elevated PARKIN levels were used to increase the likelihood of capturing signal-dependent and mechanistically relevant ubiquitylation events that would otherwise be below the level of detection due to low stoichiometry or abundance.

Using HCT116 cells expressing a haemagglutinin-derived epitope tag (HA)-fused with PARKIN (HCT116^{PARKIN}, Tier 1), we performed 34 control and QdiGly experiments (Supplementary Table 1), quantifying 4,772 non-redundant ubiquitylation sites in 1,654 proteins (Fig. 1a, b). In 18 of these experiments using Btz or CCCP/Btz for 1 h, we identified 443 diGly sites (261 proteins) with log₂(heavy:light (H:L)) ratios ≥ 1.0 (twofold change) in at least one experiment (Fig. 1b, c, Supplementary Fig. 2a–c, Supplementary Tables 1, 2), with a Pearson's correlation of 0.69 for two representative experiments (Supplementary Fig. 2e). Sixteen additional Tier 1 experiments with depolarization times up to 10 h identified 537 non-redundant diGly sites (304 proteins), including 192 diGly sites (138 proteins) also identified at 1 h (Supplementary Fig. 2d). Comparison of log₂(H:L) ratios for 48 Tier 1 sites in 36 proteins at 1 and 8 h post CCCP treatment from parallel experiments revealed persistent or increased ubiquitylation for 34 sites when proteasome activity is blocked (Supplementary Fig. 2f).

To systematically determine PARKIN dependence and compare targets in distinct cell lines, we performed 6 QdiGly experiments in HeLa versus HeLa^{PARKIN} cells (Tier 2) with CCCP/Btz (1 h) and reverse SILAC labelling (Fig. 1a, Supplementary Tables 1, 2). As expected^{4–7}, CCCP-dependent MFN2 poly-ubiquitylation in HeLa cells required PARKIN expression (Supplementary Fig. 2a). We identified 582 PARKIN-dependent diGly peptides (303 proteins) (Fig. 1b), with a Pearson's correlation of 0.88 for duplicate samples (Supplementary Fig. 2e) and significant overlap across biological triplicates (Fig. 1d). Importantly, 165 diGly sites (99 proteins) were common to Tiers 1 and 2 with 1 h of depolarization (Fig. 1e). This increased to 191 sites (144 proteins) when all Tier 1 and Tier 2 data were compared

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

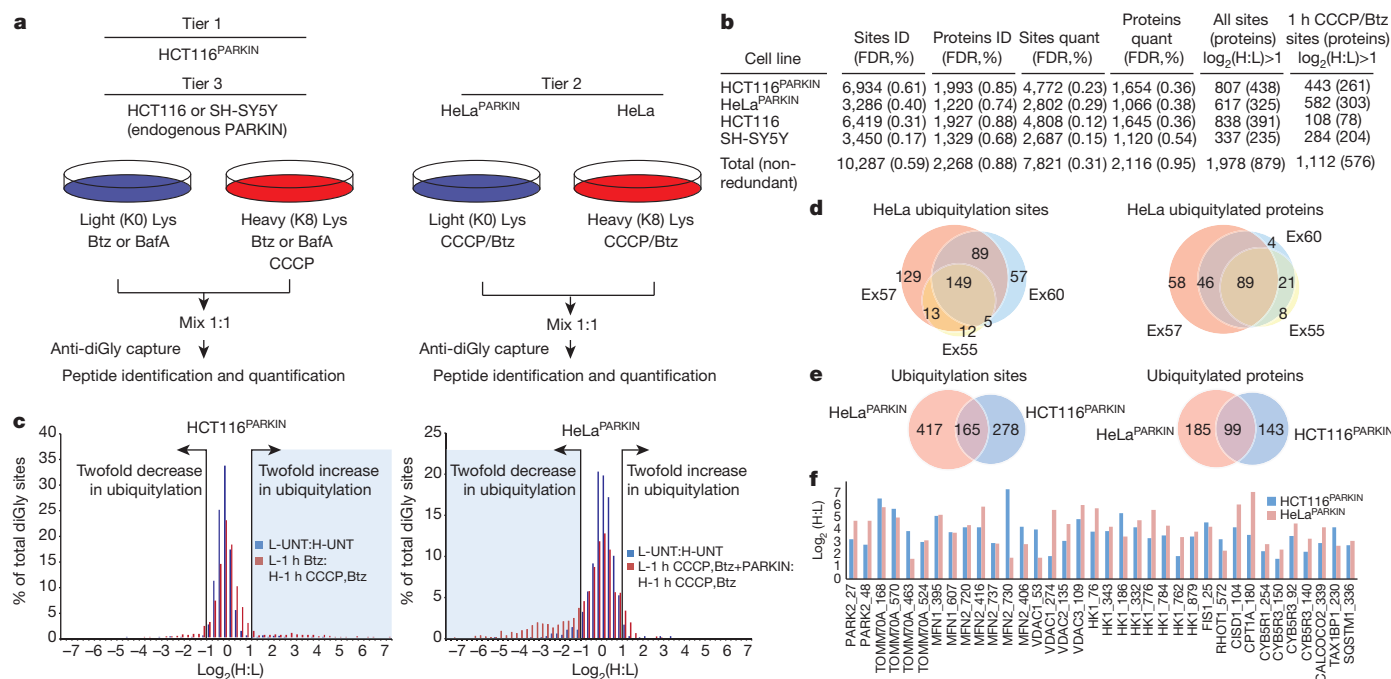


Figure 1 | QdiGly proteomics for PARKIN-dependent ubiquitylation.

a, Proteomics workflow. **b**, diGly sites identified and quantified across 73 experiments. FDR, false discovery rate. **c**, Log₂(H:L) plots for quantified diGly peptides for HCT116^{PARKIN} (experiment 17) or HeLa^{PARKIN} (experiment 57) cells (Supplementary Table 2). **d**, Overlap of ubiquitylation sites in HeLa^{PARKIN}

(Supplementary Table 2). We refer to the overlapping set of ubiquitylation sites with 1 h of depolarization and their associated proteins as Class 1 candidate PARKIN-dependent targets, whereas proteins found in both cell lines but with different sites of ubiquitylation are referred to as Class 2 (Supplementary Table 2). Whereas a twofold increase in H:L ratio was used as a threshold for regulated ubiquitylation, many diGly sites were induced 30–60-fold upon depolarization (Fig. 1f, Supplementary Table 2), indicating highly dynamic target modification via PARKIN.

Class 1 targets were enriched in mitochondrial proteins ($P < 1.76 \times 10^{-17}$), although numerous cytoplasmic proteins were also identified (Fig. 2, Supplementary Fig. 2i). In total, 60 Class 1 and 2 targets were linked with mitochondria or endoplasmic reticulum-type membranes, including 36 MOM proteins. Candidate cytoplasmic targets included proteasome subunits, the VCP/p97 ATPase, and proposed autophagy adaptors (SQSTM1, CALCOCO2 (also known as NDP52), and TAX1BP1).

We next performed 22 control and QdiGly profiling experiments using HCT116 or neuronal SH-SY5Y cells (Tier 3, Fig. 1a, b, Supplementary Tables 1, 2), which show primarily mono-ubiquitylated forms of MFN2 ubiquitylation upon depolarization, apparently due to low levels of endogenous PARKIN (Supplementary Fig. 2a). Cumulatively, 838 diGly sites (391 proteins) in HCT116 and 337 diGly sites (235 proteins) in SH-SY5Y had log₂(H:L) ≥ 1.0 (Fig. 1b), with extensive overlap across Tier 1 and 2 data sets (Supplementary Fig. 2g) and biological triplicates in SH-SY5Y (1 h/CCCP) (Supplementary Fig. 2h). Fifteen sites in 12 Class 1 or 2 targets were found in both cell lines, and 29 and 27 CCCP-dependent diGly sites in 27 and 17 Class 1 or 2 proteins were detected in SH-SY5Y or HCT116 cells, respectively (Fig. 2, Supplementary Tables 1, 2). Additionally, 124 Tier 3 sites were identified in 29 Class 1 or 2 proteins that were distinct from Class 1 sites. Thus, many PARKIN-dependent diGly targets were confirmed at endogenous PARKIN levels. Ubiquitylation of several candidate substrates was demonstrated by immunoblotting, and ubiquitylation was reduced upon small-interfering RNA (siRNA)-mediated depletion of PARKIN for several targets tested (Supplementary Fig. 3a–d).

biological triplicates (1 h CCCP + Btz) (Supplementary Tables 1, 2). Ex, experiment. **e**, Ubiquitylation site and protein overlap between all HCT116^{PARKIN} and HeLa^{PARKIN} experiments treated with CCCP and Btz for 1 h. **f**, log₂(H:L) ratios for selected diGly sites from HCT116^{PARKIN} (Ex17) and HeLa^{PARKIN} (Ex57) (1 h CCCP + Btz).

To complement QdiGly proteomics, we used anti-HA affinity purification–mass spectrometry (AP–MS)²² to identify depolarization-dependent high-confidence candidate HA–PARKIN-interacting proteins (HCIPs) in duplicate experiments in 293T^{PARKIN} cells (Fig. 3a, Supplementary Fig. 4a, Supplementary Table 3), with further validation in HeLa^{PARKIN} cells (Supplementary Fig. 4b, Supplementary Table 4) (see Methods). Because PARKIN action is dynamic, we examined several time points (1, 4 or 8 h) after depolarization in the presence or absence of Btz or BafA. To allow a semiquantitative assessment of interactions, we compared bait-normalized average assembled peptide spectral matches (APSMs) for proteins that were HCIPs under at least one condition examined unless otherwise noted (Fig. 3a, Supplementary Fig. 4a, Supplementary Tables 3, 4). Without depolarization, no HCIPs were common to both wild-type PARKIN biological duplicates, consistent with PARKIN auto-inhibition before activation²³ (Fig. 3a, Supplementary Text). Remarkably, upon depolarization, four classes of HCIPs were identified in at least one condition examined (Fig. 3a, b): (1) 20 MOM proteins, including MFN1/2, RHOT1/2 and VDACS, the fission protein FIS1 and its interacting protein TBC1D15, and the translocase components TOMM20 and TOMM70; (2) the autophagy adaptors SQSTM1 (also known as p62), CALCOCO2 and TAX1BP1; (3) subunits of the catalytic and regulatory particles of the proteasome; and (4) the VCP (also known as p97) ATPase implicated in MFN1 turnover⁷. Nineteen of the 26 mitochondrial and autophagy proteins and 39 of 40 proteasome subunits found in 293T cells were also identified in at least one condition examined in HeLa cells (Fig. 3b, Supplementary Fig. 4b). Depolarization-dependent interaction of Flag–PARKIN with selected endogenous HCIPs was verified by anti-Flag immunoprecipitation and immunoblotting (Supplementary Fig. 5). Notably, 21 of the 27 mitochondrial and autophagy receptor proteins found associated with PARKIN, as well as eight subunits of the regulatory particle of the proteasome and VCP, were also identified as Class 1 or 2 PARKIN-dependent ubiquitylation targets (Figs 2, 3b, Supplementary Table 2).

PARKIN activation is thought to reverse auto-inhibition by its amino-terminal ubiquitin-like domain (UBL), which has been proposed to

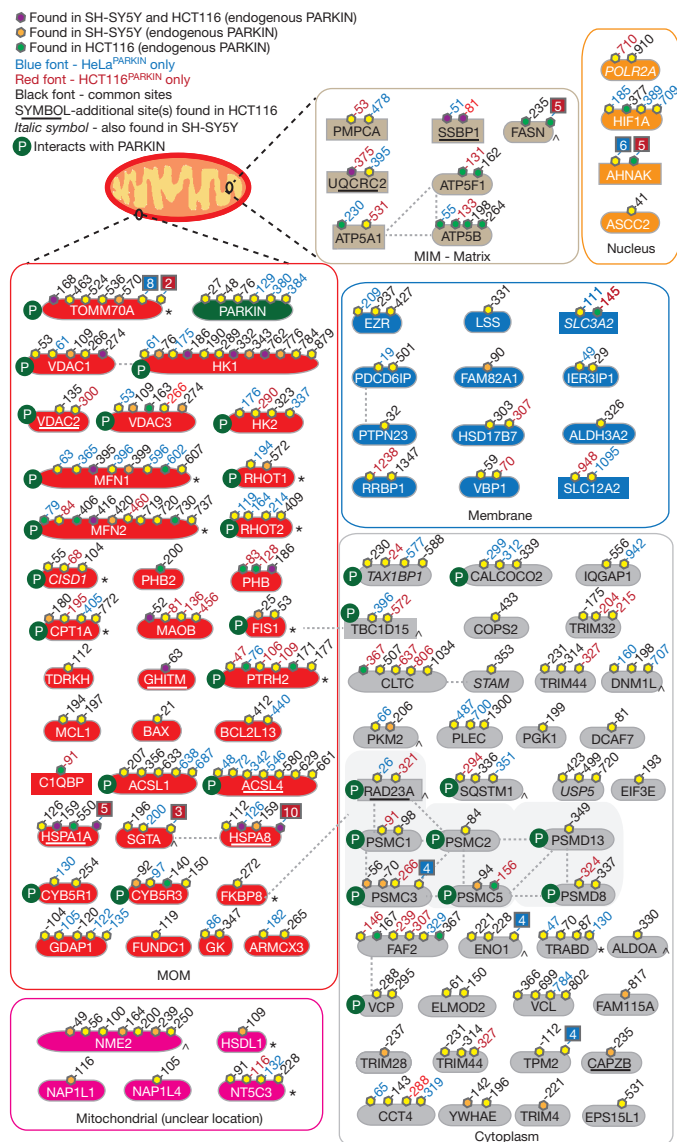
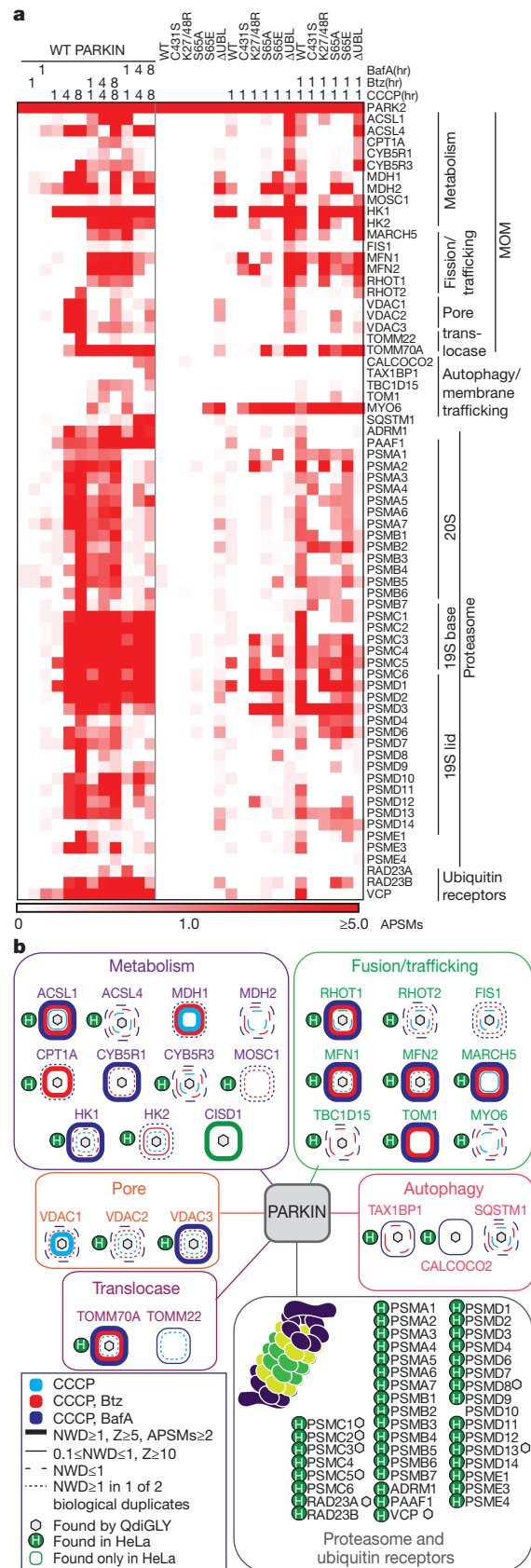


Figure 2 | PARKIN-dependent ubiquitylation sites revealed by QdiGly proteomics. Class 1 sites are in black font. Additional sites found in Class 1 proteins are in red (HCT116^{PARKIN}) or blue font (HeLa^{PARKIN}). Site overlap in HCT116 and/or SH-SY5Y: magenta, orange, and green octagons. Dotted lines: interacting proteins. Rectangles represent Class 2 substrates. Red or blue boxes refer to additional sites identified in either HCT116^{PARKIN} and HeLa^{PARKIN} cells (Supplementary Table 2). MIM, mitochondrial inner membrane. * and ^, protein levels decrease or increase, respectively, upon depolarization⁴.

Figure 3 | PARKIN associates with mitochondrial proteins and the proteasome in response to depolarization. **a**, Heat map of HCIPs (represented by APSMs) for HA-Flag-PARKIN and mutants in response to depolarization in 293T cells, with or without Btz or BafA. Proteins indicated had weighted and normalized *D*-scores ≥ 1.0 , *Z*-score ≥ 5 , and APSMs ≥ 2 unless otherwise noted (see Methods). **b**, Summary of PARKIN-interacting proteins in 293T and HeLa and integration with diGly sites.



sterically hinder ubiquitin-conjugating (E2) enzyme interaction with the catalytic centre^{20,23}. We explored the role of these regions in MOM recruitment, MFN2 ubiquitylation, substrate binding and proteasome association by using five mutants (Supplementary Fig. 6a): (1) Δ UBL, lacking the UBL domain (residues 1–80), (2) K27/48R, which blocks ubiquitylation of two regulated diGly sites in the UBL (Fig. 2), (3) S65A and S65E, which block PINK1-dependent phosphorylation of the UBL domain implicated in PARKIN activation *in vitro*¹⁸, and (4) C431S, removing the active site Cys in RING2. The Parkinson's disease patient mutation PARKIN(C431F) fails to localize to mitochondria in response to depolarization^{15,20}. Consistent with this, PARKIN(C431S) failed to localize to mitochondria, efficiently poly-ubiquitylate MFN2, or interact robustly with MOM substrates in response to depolarization (Fig. 3a, Supplementary Fig. 6a–d). In contrast, K27/K48R, S65A and S65E mutants showed near wild-type activity for localization, MFN2 ubiquitylation and interaction with MOM proteins at 1 h post depolarization (Fig. 3a, Supplementary Fig. 6a–d, Supplementary Tables 3, 6), although we cannot rule out a kinetic effect of these mutations. PARKIN(Δ UBL) displayed readily detectable association with MOM substrates in the absence of depolarization, consistent with some loss of auto-inhibition²², but nevertheless showed enhanced substrate association upon depolarization, indicating residual PINK1 dependence despite the absence of the UBL and S65. However, MFN2 ubiquitylation was reduced compared with wild-type PARKIN, suggesting reduced ubiquitylation activity. PARKIN(Δ UBL) also lacked a robust increase in proteasomal binding (Fig. 3a).

Our data indicate that PARKIN interacts with and promotes the site-specific ubiquitylation of numerous mitochondrial and cytoplasmic proteins, thereby extending previous studies examining abundance of mitochondrially enriched proteins^{1,4}. To place candidate PARKIN-dependent ubiquitination targets into a structural and functional framework, we have created an interactive resource that integrates dynamic ubiquitylation data with available structural information (Supplementary Fig. 7b–e and Methods). We explored the spatial relationship between protein localization, structure and conservation for 90 Class 1 sites on 36 candidate PARKIN substrates (Fig. 4a–d, Supplementary Fig. 7a, Supplementary Table 5). We found that 90% of ubiquitylated lysine residues are conserved in mouse (*Mus musculus*), 78% in zebrafish (*Danio rerio*), and 51% in the 29 conserved orthologues in *D. melanogaster*, which harbours an active PARKIN–PINK1 pathway⁶. Moreover, 38% of sites were conserved in all three species, and 22 of 36 proteins have at least one site conserved from human to *D. melanogaster* (Fig. 4a–d, Supplementary Fig. 7a). DiGly sites were observed on multiple structural elements (α -helices, β -sheets and loops) and no obvious sequence motif was identified by MOTIFX, indicating that PARKIN specificity is driven primarily by substrate recruitment and/or proximity rather than specific target sequences.

Of the 17 MOM targets whose structures or membrane orientations are defined, all identified ubiquitylation sites are presented on

the cytoplasmic surface (Fig. 4c, d), consistent with the idea that PARKIN re-sculpts the MOM proteome via its cytoplasmic recruitment. Fifty-nine percent of Class 1 and 2 targets had one or two sites of ubiquitylation, whereas 41% of targets had 3–15 sites that were PARKIN-dependent on the basis of Tier 2 analysis (Supplementary Fig. 7f). Of the eight Class 1 and Class 2 sites identified in MFN1, all but two are located in helical motifs flanking the two carboxy-terminal membrane-spanning regions, indicating localized ubiquitylation (Fig. 2 and Fig. 4d). We detected 14 Class 1 and 2 sites in HK1, but unlike MFN1, these sites were comparatively delocalized across both globular domains in HK1 (Fig. 4c). Similarly, TOMM70 was decorated extensively over its 10 cytoplasmic TPR motifs (Figs 2, 4d). Although we cannot exclude the possibility that some ubiquitylation events occur downstream of PARKIN activation, the fact that we find many of the targets as PARKIN-interacting proteins (Fig. 3b) is consistent with direct ubiquitylation (see Supplementary Text). An understanding of how PARKIN ubiquitylates topologically diverse proteins and domains within proteins requires further study but will be facilitated by the site-specificity data reported here.

PARKIN seems to promote rapid turnover of some targets but not others. In total, 12 of the Class 1 and 2 proteins identified here (see Fig. 2), like MFN1/2 and RHOT1/2^{1,4,7,10,11}, probably undergo proteasomal turnover, including C1QBP1 (Supplementary Fig. 3c, d), FIS1 and CISD1, as their total levels are also decreased upon PARKIN activation^{1,4}. In contrast, previous studies⁴ have identified 50 preferentially cytoplasmic proteins or protein complexes that are enriched in mitochondria in response to depolarization. Consistent with this, we identified 11 of these proteins or components of protein complexes as Class 1 or 2 targets (Fig. 2 and Supplementary Fig. 7g), possibly reflecting an under-appreciated dynamic recruitment process affecting mitochondrial homeostasis. These include the autophagy adaptor p62, previously linked with mitochondrial clustering after damage²⁴, as well as TBC1D15 and DNML1, which are known to interact with FIS1 to regulate fission–fusion cycles²⁵. In addition to p62, we also identified candidate autophagy receptors, CALCOCO2 and TAX1BP1, which are ubiquitylated upon depolarization and associate with PARKIN in the presence of BafA (Figs 2, 3). Because CALCOCO2 has been shown to target bacteria for autophagy²⁶, it is an attractive candidate for involvement in mitophagy. We speculate that these related autophagy adaptor proteins may be dynamically recruited to the MOM and may link the depolarization response to the recruitment of autophagy machinery. Additionally, it is possible that additional cytoplasmic proteins identified here as PARKIN-dependent targets are transiently recruited to mitochondria but below the level of detection in previous studies. It is also likely that ubiquitylation of PARKIN targets alters their enzymatic properties or functions before their turnover by the proteasome or autophagy. Potential PARKIN-regulated functions include: (1) fission–fusion cycles (FIS1, TBC1D15), (2) small molecule transport (VDACs), (3) apoptosis (FKBP8, MCL1, BAX), (4)

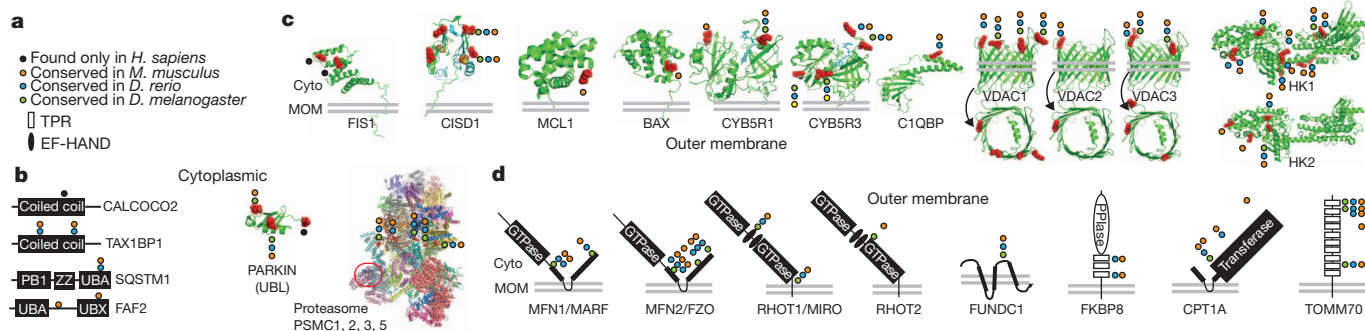


Figure 4 | Structural anatomy and conservation of PARKIN-dependent diGly sites. Structures or domain schematics are shown for Class 1 depolarization and PARKIN-dependent diGly sites (PDB codes, Supplementary Table 5). **a**, Legend for colour-coded circles that indicate the

conservation of Lys at homologous positions in *M. musculus*, *D. rerio* and *D. melanogaster* (Supplementary Table 5). For structures, regulated sites are shown in red space-filled models. **b**, Cytoplasmic proteins. RPN10, red circle. **c**, MOM proteins with structures. **d**, MOM proteins without structures.

iron-sulphur shuttling (CISD1), (5) protein translocation (TOMM70) or (6) proteasome assembly or activity (PSMC subunits) (see Supplementary Text)^{25,27}.

This work provides systematic target identification and ubiquitylation site-specificity for PARKIN, thereby revealing the diversity and complexity of PARKIN function on the MOM and in the cytoplasm. This resource will enable the development of methods that detect specific diGly sites in PARKIN targets *in situ*, thus enabling a deeper understanding of how PARKIN controls mitochondrial homeostasis in tissues relevant to Parkinson's disease.

METHODS SUMMARY

Cells were grown in Lys(K)-free DMEM/dialysed fetal bovine serum, with light (K0) or heavy (K8) Lys (50 µg ml⁻¹)¹², and treated with Btz (1 µM), CCCP (10 µM), and/or BafA (50 nM). Mixed cells were lysed in 8 M urea, 75 mM NaCl, 50 mM Tris HCl (pH 8.2), and tryptic/LysC peptides immunoprecipitated with anti-diGly-Protein A (Cell Signaling Technologies) (1 h, 4 °C), typically with 2–4 sequential immunoprecipitations (Supplementary Table 1)¹². Peptides and ubiquitylation sites were identified by liquid chromatography-tandem mass spectrometry (see Methods)¹². Extracted ion chromatograms were used for quantification and signal-to-noise measurements¹². Interaction proteomics used 293T or HeLa cells stably expressing HA–Flag–PARKIN (or mutants) and CompPASS to identify HCIPs²².

Full Methods and any associated references are available in the online version of the paper.

Received 11 November 2012; accepted 26 February 2013.

Published online 17 March; corrected online 17 April 2013 (see full-text HTML version for details).

- Narendra, D., Walker, J. E. & Youle, R. Mitochondrial quality control mediated by PINK1 and Parkin: links to parkinsonism. *Cold Spring Harb. Perspect. Biol.* **4**, a011338 (2012).
- Youle, R. J. & Narendra, D. P. Mechanisms of mitophagy. *Nature Rev. Mol. Cell Biol.* **12**, 9–14 (2011).
- Dawson, T. M. & Dawson, V. L. The role of parkin in familial and sporadic Parkinson's disease. *Mov. Disord.* **25** (Suppl 1), S32–S39 (2010).
- Chan, N. C. *et al.* Broad activation of the ubiquitin-proteasome system by Parkin is critical for mitophagy. *Hum. Mol. Genet.* **20**, 1726–1737 (2011).
- Glauser, L., Sonnay, S., Stafa, K. & Moore, D. J. Parkin promotes the ubiquitination and degradation of the mitochondrial fusion factor mitofusin 1. *J. Neurochem.* **118**, 636–645 (2011).
- Ziviani, E., Tao, R. N. & Whitworth, A. J. *Drosophila* parkin requires PINK1 for mitochondrial translocation and ubiquitinates mitofusin. *Proc. Natl Acad. Sci. USA* **107**, 5018–5023 (2010).
- Tanaka, A. *et al.* Proteasome and p97 mediate mitophagy and degradation of mitofusins induced by Parkin. *J. Cell Biol.* **191**, 1367–1380 (2010).
- Poole, A. C., Thomas, R. E., Yu, S., Vincow, E. S. & Pallanck, L. The mitochondrial fusion-promoting factor mitofusin is a substrate of the PINK1/parkin pathway. *PLoS ONE* **5**, e10054 (2010).
- Sun, Y., Vashisht, A. A., Tchiew, J., Wohlschlegel, J. A. & Dreier, L. VDACs recruit Parkin to defective mitochondria to promote mitochondrial autophagy. *J. Biol. Chem.* **287**, 40652–40660 (2012).
- Wang, X. *et al.* PINK1 and Parkin target Miro for phosphorylation and degradation to arrest mitochondrial motility. *Cell* **147**, 893–906 (2011).
- Liu, S. *et al.* Parkinson's disease-associated kinase PINK1 regulates Miro protein level and axonal transport of mitochondria. *PLoS Genet.* **8**, e1002537 (2012).
- Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).
- Wagner, S. A. *et al.* A proteome-wide, quantitative survey of *in vivo* ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell Proteomics* **10**, M111.013284 (2011).
- Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UbqH7-reactivity profile reveals Parkin and HHAR1 to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
- Joselin, A. P. *et al.* ROS-dependent regulation of Parkin and DJ-1 localization during oxidative stress in neurons. *Hum. Mol. Genet.* **21**, 4888–4903 (2012).
- Narendra, D., Tanaka, A., Suen, D. F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803 (2008).
- Lazarou, M., Jin, S. M., Kane, L. A. & Youle, R. J. Role of PINK1 binding to the TOM complex and alternate intracellular membranes in recruitment and activation of the E3 ligase Parkin. *Dev. Cell* **22**, 320–333 (2012).
- Kondapalli, C. *et al.* PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biol.* **2**, 120080 (2012).
- Sha, D., Chin, L. S. & Li, L. Phosphorylation of parkin by Parkinson disease-linked kinase PINK1 activates parkin E3 ligase function and NF-κB signaling. *Hum. Mol. Genet.* **19**, 352–363 (2010).
- Lazarou, M. *et al.* PINK1 drives Parkin self-association and HECT-like E3 activity upstream of mitochondrial binding. *J. Cell Biol.* **200**, 163–172 (2013).
- Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
- Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389–403 (2009).
- Chaugule, V. K. *et al.* Autoregulation of Parkin activity through its ubiquitin-like domain. *EMBO J.* **30**, 2853–2867 (2011).
- Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. p62/SQSTM1 is required for Parkin-induced mitochondrial clustering but not mitophagy; VDAC1 is dispensable for both. *Autophagy* **6**, 1090–1106 (2010).
- Onoue, K. *et al.* Fis1 acts as mitochondrial recruitment factor for TBC1D15 that is involved in regulation of mitochondrial morphology. *J. Cell Sci.* <http://dx.doi.org/10.1242/jcs.111211> (2012).
- von Muhlinen, N. *et al.* LC3C, bound selectively by a noncanonical LIR motif in NDP52, is required for antibacterial autophagy. *Mol. Cell* **48**, 329–342 (2012).
- Huang, P., Galloway, C. A. & Yoon, Y. Control of mitochondrial morphology through differential interactions of mitochondrial fusion and fission proteins. *PLoS ONE* **6**, e20655 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. Kim for LC-MS and for development of QdiGly profiling, R. Kunz for peptide purification, J. Lydeard, S. Hayes and A. White for proteomics, M. Comb and S. Beausoleil (Cell Signaling Technologies) for antibodies, Nikon Imaging Center (Harvard Medical School) for microscopy, and D. Finley and B. Schulman for discussions. Supported by NIH grants GM070565 and GM095567 to J.W.H., GM067945 to S.P.G., CA139885 to M.R., and the Michael J. Fox Foundation for Parkinson's Research to J.W.H.

Author Contributions S.A.S. and J.W.H. conceived the experiments. S.A.S. performed QdiGly profiling, biochemical, and interaction experiments and analysis. S.A.S., M.R. and V.G.-P. performed cell biological experiments. M.E.S. designed site visualization software. E.L.H. and S.P.G. provided proteomics software and analysis. S.A.S. and J.W.H. wrote the manuscript. All authors assisted in editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.H. (wade_harper@hms.harvard.edu).

METHODS

Sample preparation for diGly capture. HCT116, HeLa or SH-SY5Y cells (10^8) were grown in lysine-free DMEM supplemented with 10% dialysed FBS (Gibco), 2 mM L-glutamine, 1% penicillin/streptomycin (Gibco), and light (KO) lysine ($50 \mu\text{g ml}^{-1}$), which contained no heavy isotopes¹². Heavy media was the same except the light lysine was replaced with heavy (K8, 8 Da heavier) lysine (L-lysine:2HCl ($\text{U-}^{13}\text{C}_6$, 99%, $\text{U-}^{15}\text{N}_2$, 99%); Cambridge Isotopes) at the same concentration. Where indicated, cells were treated with Btz ($1 \mu\text{M}$), CCCP ($10 \mu\text{M}$), and/or BafA (50 nM) for the times indicated. Btz was a gift from Millennium Pharmaceuticals. After the indicated treatments, heavy and light cells were mixed 1:1 by cell number and lysed in 8 ml of denaturing lysis buffer (8 M urea, 50 mM Tris (pH 8.2), 75 mM NaCl, protease inhibitors (EDTA-free), Roche)¹². Samples were incubated on ice for 10 min and then sonicated with $3 \times 10 \text{ s}$ pulses. We typically obtained 30–50 mg of total protein. Lysates were digested with trypsin as described previously²⁸ with one modification. Prior to trypsin digestion, lysates were diluted 1:1 with 50 mM Tris (pH 8.2) to lower the urea concentration to 4 M and digested with $10 \text{ ng } \mu\text{l}^{-1}$ Lys-C (Wako) for 2 h at room temperature.

Immunoprecipitation of diGly containing peptides. Lyophilized peptides from 30–50 mg of digested proteins were resuspended in 1.3 ml of IAP buffer (50 mM MOPS (pH 7.4), 10 mM Na_2HPO_4 , 50 mM NaCl) and centrifuged at $14,000g$ for 5 min to remove any insoluble materials. The supernatant was incubated with anti-diGly antibody coupled to protein A agarose or acrylamide beads for 1 h at 4°C and washed with IAP buffer $3 \times$ and once with PBS as described previously¹². Peptides were eluted by treatment with $50 \mu\text{l}$ of 5% formic acid for 10 min twice. The eluted peptides were desalted using C18 stage-tip method and resuspended in 5% formic acid before mass spectrometric analysis. Lysates were subjected to immunoprecipitation sequentially two or four times, unless otherwise noted in Supplementary Table 1.

Mass spectrometry analysis of diGly peptides. Peptides were separated on $100 \mu\text{m} \times 20 \text{ cm}$ C18 reversed phase (Maccel C18 $3 \mu\text{m}$ 200 Å, The Nest Group) with a 165 min gradient of 6% to 27% acetonitrile in 0.125% formic acid²⁹. The twenty most intense peaks from each full mass spectrometry (MS) scan acquired in the Orbitrap Velos (Thermo) were selected for MS/MS (see RAW files for specific settings).

Sequest-based identification using a Human UNIPROT database followed by a target decoy-based linear discriminant analysis was used for peptide and protein identification as described^{12,30,31}. Localization of diGly sites used a modified version of the A-score algorithm³² as described¹². A-scores > 13 were considered localized. SILAC-based site quantification and signal-to-noise was performed as described previously¹² by using extracted ion chromatograms. The heavy- and light-labelled peptides from each search were subsequently combined using custom scripts. Other parameters used for database searching include: 50 p.p.m. precursor mass tolerance; 1.0 Da product ion mass tolerance; tryptic digestion with up to three missed cleavages; and variable oxidation of Met (+15.994946). For an individual peptide to be used for quantification of sites and proteins, it had to meet one of two conditions: (1) both heavy and light isotopic envelopes must be detected with signal-to-noise ratios above 5.0; and (2) one isotopic envelope (heavy or light) must have a signal-to-noise ratio above 10.0. Equal mixing ratios for heavy to light cells were confirmed by calculating the mean $\log_2(\text{H:L})$ ratios in non-diGly non-lysine containing peptides and subtracting that from the corresponding $\log_2(\text{H:L})$ diGly ratios if necessary. Keratins were removed from site lists. Contributions of peptide identifications to various protein isoforms were condensed into a single UniProt descriptor using principles of parsimony. Mass spectrometry RAW files are available upon request.

Interaction proteomics. Interaction proteomics was performed essentially as described previously, but with small modifications²². Briefly, 293T or HeLa cells were transfected with a lentiviral vector expressing HA-Flag-PARKIN (NP_004553.2) or the designated PARKIN mutants, and stable cell lines selected in puromycin. Cells from $4 \times 15 \text{ cm}$ dishes at 80% confluence were treated with CCCP ($10 \mu\text{M}$), Btz ($1 \mu\text{M}$), and/or BafA (50 nM) as indicated and after the indicated time, cells were collected and lysed in 3 ml of 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Nonidet P40, and protease inhibitors. Cleared lysates were filtered through $0.45 \mu\text{m}$ spin filters (Millipore Ultrafree-CL) and immunoprecipitated with $30 \mu\text{l}$ anti-HA resin (Sigma). Complexes were washed $4 \times$ with lysis buffer, exchanged into PBS for a further three washes, eluted with HA peptide, reductively carboxymethylated, and precipitated with 10% trichloroacetic acid (TCA). TCA-precipitated proteins were trypsinized, purified with Empore C18 extraction media (3 M), and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) with a LTQ-Velos linear ion trap mass spectrometer (Thermo) with an 18 cm^3 125 μm (ID) C18 column and a 50 min 8–26% acetonitrile gradient. All AP-MS experiments in 293T cells were performed in biological duplicates and for each biological experiment, complexes were analysed twice by LC-MS to generate technical duplicates. AP-MS experiments in

HeLa cells were performed on a single IP but with technical duplicates. Spectra were searched with Sequest against a target-decoy human tryptic UniProt-based peptide database, and these results were loaded into the Comparative Proteomics Analysis Software Suite (CompPASS) to identify high confidence candidate interacting proteins (HCIPs)²². Here, a statistics table derived from analogous AP-MS data for 166 unrelated proteins was used to determine weighted and normalized D-scores (WD^{N} -score) as well as Z-scores based on spectral counts. The D-score measures the reproducibility, abundance and frequency of individual proteins detected in each individual analysis.

To identify PARKIN-associated proteins, we filtered proteins at a 2% false discovery rate for those with a WD^{N} -score ≥ 1.0 , Z-score ≥ 5 , and average assembled peptide spectral matches (APSMs) ≥ 2 in both biological duplicates. In addition to the initially identified proteins, APSMs for interactors with WD^{N} -scores greater than 1 under at least one condition examined, or with a Z-score greater than 10 in both biological duplicates, were plotted in heat map form using Multi-experiment Viewer software (MeV). Due to space limitations, a subset of proteins which had a WD^{N} -score ≥ 1.0 in at least one experiment were omitted from Fig. 3 but are contained in the heat maps in Supplementary Fig. 4 as well as Supplementary Tables 3 and 4. TAX1BP1 that was identified in PARKIN immunoprecipitates and also found to be ubiquitinated but did not pass the stringent cut-off for passing our scoring scheme was also displayed in the heat map. We note that whereas several proteins passed the WD^{N} -score cut-off for PARKIN-associated proteins in the absence of depolarization in individual experiments, there was essentially no overlap within biological duplicates (Supplementary Table 4). In addition, for proteins with a maximum of 1–3 spectral counts, it is possible that those proteins were missed under some conditions due to stochastic sampling of low abundance proteins by LC-MS (for example, CYB5R3, TOM22, and RHOT1 with wild-type PARKIN at 8 h post-CCCP in the presence versus the absence of BafA, Fig. 3a). The following proteins were identified in only 1 biological duplicate as being associated with PARKIN but were included in the heat map (Fig. 3a) as they are also present as Class 1 diGly targets: SQSTM1, TAX1BP1, TBC1D15, MYO6, VCP, RHOT2, HK2, MDH2, CYB5R3 and ACSL4.

Immunological methods and microscopy. To validate interactions between PARKIN and candidate interacting proteins (Supplementary Fig. 3), 293T cells stably expressing HA-Flag-PARKIN were either left untreated or treated with CCCP ($10 \mu\text{M}$) or CCCP ($10 \mu\text{M}$) and Btz ($1 \mu\text{M}$) for 1 h. Extracts (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.5% Nonidet P40, and protease inhibitors) from cells were subjected to immunoprecipitation with anti-Flag resin (M2 agarose; Sigma), and washed complexes subjected to immunoblotting with the indicated antibodies. To examine MFN2 ubiquitylation, extracts from cells lysed in denaturing lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris (pH 8.0), and protease inhibitors) expressing the indicated HA-Flag-PARKIN mutant proteins were separated on a 4–12% gradient SDS-polyacrylamide gel (PAGE) and blotted with anti-HA or anti-MFN2. To examine PARKIN levels in the cell lines used (Supplementary Fig. 1), cells were lysed in denaturing lysis buffer and protease inhibitors and extracts subjected to SDS-PAGE and immunoblotting. To examine ubiquitylation of candidate PARKIN substrates by immunoblotting, HCT116 cell lines stably expressing the candidate substrates C-terminally tagged with HA-Flag epitopes were transfected with either control siRNA or an siRNA targeting PARKIN (1: 5'-GGAGUGCAGUGCCGUAUUU-3', 2: 5'-GUAAGAAGCGUACCAUGA-3', 3: 5'-GGAGCCAUCGCCUAUGAAA-3', 4: 5'-GUAAGAAGCGUACCAUGA-3', Thermo Scientific). After 72 h, cells were treated with CCCP ($10 \mu\text{M}$) with or without Btz to block proteasomal turnover of targets. At the indicated times, cells were lysed in 8 M urea and subjected to immunoblotting using anti-HA antibodies. Identical gels were run and probed with anti-PCNA as a loading control. The antibodies used are as follows: anti-TOMM70, (Proteintech, 14528-1); anti-MFN2, (Epitomics, 3272-1 or Sigma, WH0009927M3); anti-ADRM1, (Enzo Scientific, BML-PW9910); anti-Rpn10, (Enzo Scientific, BML-PW9250); anti-HK1, (Cell Signaling, C35C4); anti-PSMD2, (Affinity BioReagents, PA1-964); anti-PARKIN, (PRK8, Santa Cruz, sc-32282).

To examine the localization of PARKIN and various PARKIN mutants, HeLa cells stably expressing HA-Flag-tagged proteins were plated on No.1 coverslips, treated with CCCP ($10 \mu\text{M}$, 1 h), and fixed with 4% paraformaldehyde before immunofluorescence using anti-HA to detect PARKIN proteins and anti-TOMM20 to detect mitochondria. All images were collected with a Yokogawa CSU-X1 spinning disk confocal on a Nikon Ti-E inverted microscope equipped with $\times 100$ Plan Apo numerical aperture 1.4 objective lens. HA-PARKIN fluorescence was excited with the 491 nm line (selected with an AOTF) from Spectral Applied Precision LMM-7 solid-state laser launch. Emission was collected with a quad band pass polychroic mirror (Semrock) and a Chroma ET525/50m emissions filter. TOMM20 fluorescence was excited with the 561 nm line from the LMM-7 launch, and emission collected with the Semrock polychroic and a

Chroma ET620/60m emission filter. Images were acquired with a Hamamatsu ORCA-AG cooled CCD camera controlled with MetaMorph 7 software. Nine *z*-series optical sections were collected with a step size of 0.2 μm , using the internal Nikon Ti-E focus motor. *z*-series were deconvolved using AutoQuant blind deconvolution software, and are displayed as maximum *z*-projections. Gamma and brightness were adjusted on displayed images (identically for compared image sets) and percent colocalization was calculated by thresholding for signal at least 4 standard deviations above background in a single *z* slice (identically for compared image sets) using MetaMorph 7 software (Supplementary Table 6).

Structural analysis of ubiquitylation sites and web portal. Using the UniProt identifier for each protein identified, the corresponding PDB file(s) (if available) were determined using data from the UniProt to PDB cross-reference table (<http://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>). The amino acid position in the structure corresponding to each ubiquitylation site was calculated based on their relative position in the structure using the information provided by the cross reference table. Data for each experiment was stored in MySQL tables, which are used when accessing the data via the web portal. For rendering of the structural information via the web portal, Jmol: an open-source Java viewer for chemical structures in 3D

(<http://www.jmol.org/>) was used. Access to the web portal is available at hms.harvard.edu.

Gene ontology. Gene ontology analysis was performed using DAVID³³. Additional sub-cellular localization parameters as designated in Fig. 2 were performed manually using MITOCARTA, Human Protein Atlas, and GenBank.

28. Villén, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **3**, 1630–1638 (2008).
29. Haas, W. *et al.* Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **5**, 1326–1337 (2006).
30. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
31. Eng, J. K., McCormack, A. L. & Yates, J. R. III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
32. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnol.* **24**, 1285–1292 (2006).
33. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).

A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly

Daniele Canzio^{1,2}, Maofu Liao¹, Nariman Naber¹, Edward Pate³, Adam Larson^{1,4}, Shenping Wu¹, Diana B. Marina^{1,4}, Jennifer F. Garcia^{1,4}, Hiten D. Madhani¹, Roger Cooke¹, Peter Schuck⁵, Yifan Cheng¹ & Geeta J. Narlikar¹

A hallmark of histone H3 lysine 9 (H3K9)-methylated heterochromatin, conserved from the fission yeast *Schizosaccharomyces pombe* to humans, is its ability to spread to adjacent genomic regions^{1–6}. Central to heterochromatin spread is heterochromatin protein 1 (HP1), which recognizes H3K9-methylated chromatin, oligomerizes and forms a versatile platform that participates in diverse nuclear functions, ranging from gene silencing to chromosome segregation^{1–6}. How HP1 proteins assemble on methylated nucleosomal templates and how the HP1–nucleosome complex achieves functional versatility remain poorly understood. Here we show that binding of the key *S. pombe* HP1 protein, Swi6, to methylated nucleosomes drives a switch from an auto-inhibited state to a spreading-competent state. In the auto-inhibited state, a histone-mimic sequence in one Swi6 monomer blocks methyl-mark recognition by the chromodomain of another monomer. Auto-inhibition is relieved by recognition of two template features, the H3K9 methyl mark and nucleosomal DNA. Cryo-electron-microscopy-based reconstruction of the Swi6–nucleosome complex provides the overall architecture of the spreading-competent state in which two unbound chromodomain sticky ends appear exposed. Disruption of the switch between the auto-inhibited and spreading-competent states disrupts heterochromatin assembly and gene silencing *in vivo*. These findings are reminiscent of other conditionally activated polymerization processes, such as actin nucleation, and open up a new class of regulatory mechanisms that operate on chromatin *in vivo*.

HP1 has two structured domains, a chromodomain (CD) and a chromoshadow domain (CSD), connected by an unstructured hinge region (Fig. 1a). The CD recognizes the trimethylated H3K9 (H3K9me3) mark^{7–9}, whereas the CSD can homodimerize^{10–12} and binds specific protein sequences^{13,14}. The hinge is implicated in sequence-independent RNA and DNA binding^{15,16}. Here we investigate how the key *S. pombe* HP1 protein Swi6 uses its different domains to create a regulatable HP1–chromatin complex.

It is proposed that heterochromatin spread relies on the ability of HP1 proteins to self-associate on chromatin^{1,5}. To understand how Swi6 self-association is regulated by chromatin, we first characterized the individual oligomerization equilibria in the absence of nucleosomes using analytical ultracentrifugation (AUC). Previous work has characterized at least three Swi6 oligomeric states: a monomer, a dimer mediated by CSD–CSD interactions and higher-order oligomers mediated by CD–CD interactions between dimers^{10,12,15,17,18}. Analysis of our AUC data best describes the system as a two-step self-association process: a tight association of two Swi6 monomers with an affinity constant, $K_{\text{obs}}^{\text{dim}}$, ($1/K_{\text{obs}}^{\text{dim}} < 1 \text{ nM}$, at 8 °C), followed by progressive self-association of Swi6 dimers with an identical chain-elongation affinity constant, $K_{\text{obs}}^{\text{iso}}$, ($1/K_{\text{obs}}^{\text{iso}} \sim 70 \text{ } \mu\text{M}$, at 8 °C) (dim indicates dimerization, iso indicates isodesmic and obs indicates observed; Fig. 1b–d and Supplementary Figs 1, 2 and 3). This second step, also known as

isodesmic self-association, is analogous to the self-association of tubulin dimers¹⁹.

We next tested whether the most distinguishing feature of the chromatin template, the H3K9 methyl mark, increases Swi6 oligomerization when it occupies the CD. An increase in oligomerization would be reflected by an increase in the overall weighted average sedimentation coefficient, s_{w} , of Swi6 as a function of H3K9me3 peptide (Fig. 1e). In contrast to our simplest expectation, addition of the methylated peptide reduced the value of s_{w} , implying that Swi6 self-association is inhibited by the methylated H3 tail peptide (Fig. 1e). This result suggested that the methylated H3 tail peptide and the CD–CD interface may compete for the same site. We noticed that the CD of Swi6 contains a sequence (92-Ala-Arg-Lys-Gly-Gly-Gly-97) on a loop that resembles the amino acid sequence of the H3 tail surrounding the Lys9 position (7-Ala-Arg-Lys-Ser-Thr-Gly-12) (Fig. 1f). Interestingly, although the Swi6 sequence degenerates in higher organisms to just the lysine and proximal glycine (Fig. 1f), in human HP1 isoforms the lysine shows post-translational modifications found on H3K9, such as monomethylation and acetylation²⁰. We therefore proposed that the ARK loop from the CD of one Swi6 could occupy the H3K9-binding site in another CD to mediate CD–CD self-association in solution (Fig. 1g). This is reminiscent of observations that the HP1 CD can bind ARK-containing motifs in histone H1 and G9a proteins^{21,22}.

To test this model, we investigated the effects of replacing the Arg 93 and the Lys 94 residues with alanines (Swi6^{LoopX}; Fig. 1g and Supplementary Table 1) on oligomerization. As predicted by the model, the Swi6^{LoopX} mutant showed a small but reproducible (threefold) decrease in the isodesmic affinity constant ($K_{\text{obs}}^{\text{iso}}$) (Fig. 2a and Supplementary Fig. 4). Interestingly, we noticed a substantially larger (14-fold) reduction in the association constant for dimerization ($K_{\text{obs}}^{\text{dim}}$) (Fig. 2b and Supplementary Fig. 4). Thus, in addition to the previously identified CSD–CSD interface, the ARK loop–CD interaction also participates in stabilizing a Swi6 dimer. We further found that Swi6^{LoopX} binds tail peptides approximately sixfold more strongly than wild-type Swi6 (Swi6^{WT}) (Fig. 2d), and that Swi6 dimerization is weakened with saturating methylated H3 tail peptide (Supplementary Fig. 4). These results indicate that the ARK loop–CD interaction is mutually exclusive with H3 tail binding.

The above data suggest that a Swi6 dimer can exist in at least two states: a closed state in which the ARK loop engages the CD of its partner Swi6 and an open state in which the ARK loop–CD interaction is broken (Fig. 2c). Self-association of dimers then consists of: (1) a conformational step between closed and open states (K^{conf}); and (2) a self-association step between dimers in the open state (K^{oligo}). For Swi6^{WT} the measured isodesmic association step ($K_{\text{obs}}^{\text{iso}}$) is a product of K^{conf} and K^{oligo} (Fig. 2c). In Swi6^{LoopX} the effect on dimerization masks the destabilizing effect of the loop mutations on the actual oligomerization step (K^{oligo}) (Fig. 2c).

¹Department of Biochemistry and Biophysics, University of California San Francisco, California 94158, USA. ²Chemistry and Chemical Biology Graduate Program University of California San Francisco, California 94158, USA. ³Voiland School of Chemical Engineering and Bioengineering, Washington State University, Pullman, Washington 99164, USA. ⁴Tetrad Graduate Program University of California San Francisco, California 94158, USA. ⁵National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, Maryland 20892, USA.

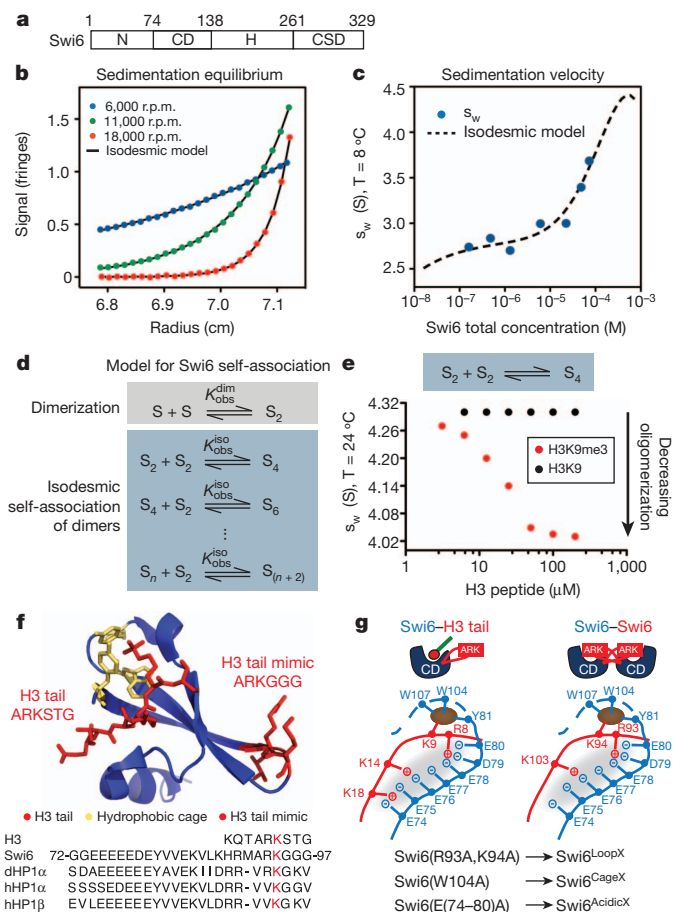


Figure 1 | Dissecting Swi6 self-association equilibria. **a**, Swi6 domains. H, hinge; N, N-terminal region. **b**, Sedimentation equilibrium AUC analysis of Swi6^{WT} self-association. Interference profiles at different rotor speeds are shown. Every twentieth point is shown. **c**, Sedimentation velocity AUC analysis of Swi6^{WT} self-association. T , temperature. For **b** and **c**, best fits for global analysis using an isodesmic self-association model are shown. **d**, Model of Swi6 self-association. S , Swi6 monomer. **e**, $[\text{Swi6}^{\text{WT}}] = 20 \mu\text{M}$. Sedimentation coefficients: dimer (S_2) = ~ 4 S; tetramer (S_4) = ~ 5.2 S. **f**, Top, Swi6 CD modelled on *Drosophila* HP1 CD with H3K9me3 peptide (Protein Data Bank (PDB) code, 1KNE). Bottom, H3 tail (amino acids 4–14) and CD loop regions of Swi6 (amino acids 72–97), *Drosophila* (d) HP1 α and human (h) HP1 α and HP1 β . Conserved lysine in red. **g**, Top, models for CD–CD and CD–ARK loop interactions. H3 tail, green line; methylation, red circle. Middle, schematic of Swi6–H3 tail interactions (left) and of hypothetical CD–CD interactions (right). Grey oval, region of negative charge; brown oval, π -cation interactions. Bottom, mutants used.

To investigate the extent of similarity between the ARK loop–CD interaction and the H3–CD interaction, we used two additional mutants. The first is Swi6^{CageX}, in which an aromatic cage residue important for H3K9me3 binding⁸ is mutated to alanine (Fig. 1g and Supplementary Table 1). The second is Swi6^{AcidicX}, in which an acidic stretch amino-terminal to the first aromatic cage residue of the CD is mutated to alanines (Fig. 1g and Supplementary Table 1; rationale in Supplementary Discussion). Both mutants show reduced binding to H3K9me3 peptides (Fig. 2d). These mutants also destabilize Swi6 oligomerization and dimerization (Fig. 2a, b and Supplementary Fig. 4), suggesting that similar interactions are involved in the H3–CD and ARK loop–CD interfaces (Fig. 1g).

We next used electron paramagnetic resonance (EPR) spectroscopy to ask whether disruption of the ARK loop–CD interface or binding of the H3K9me3 tail makes the loop more mobile by stabilizing the open conformation (Fig. 2c, e). Changes in the mobility of a site-specifically attached spin probe can give well-defined changes in its EPR spectrum²³. We mutated all three native cysteines in Swi6 to serines (Swi6³⁵,

Supplementary Table 1), mutated the Gly 95 residue on the loop to a cysteine, and then modified it with a maleimide spin probe (Swi6^{probe}, Supplementary Table 1). Mutating the native cysteines destabilized oligomerization, H3 peptide binding and nucleosome binding (Supplementary Fig. 5), but the mutants still showed substantial discrimination for the H3K9 methyl mark (Supplementary Fig. 5).

Two spectral components were observed for Swi6^{probe-WT}, one with higher mobility and one with reduced mobility. Deconvolution of the two components gave the fraction of probes that are immobilized. In parallel, AUC experiments confirmed the oligomeric state of the protein. For the Swi6^{probe-WT} protein, $\sim 35\%$ of the probes were immobile (Fig. 2f). Compared to Swi6^{probe-WT}, the fraction of immobile probes decreased in Swi6^{probe-LoopX}, Swi6^{probe-AcidicX} and Swi6^{probe-DimerX} (L151D; Supplementary Table 1), which disrupts CSD–CSD dimerization and increases monomeric Swi6 (ref. 18) (Fig. 2f and Supplementary Fig. 5). The values obtained for the mutants relative to wild type are consistent with our thermodynamic characterization (Fig. 2a, b). Further, as predicted by the model (Fig. 2c, e), addition of the H3K9me3 peptide decreased the fraction of immobile probes. The H3K9me3 peptide was ~ 100 -fold better at decreasing the immobile probe fraction compared to both H3K9 and H3K4me3 peptides, indicating that the effect was specific for the H3K9me3 mark (Fig. 2g; additional mutants in Supplementary Fig. 5).

To investigate the global structure of Swi6 dimers, we used negative-stain electron microscopy. To increase the mass for visualization by electron microscopy, and to identify the N terminus of Swi6, we fused a cyan fluorescent protein (CFP) molecule to the N terminus of Swi6 (Fig. 3a and Supplementary Fig. 6). The CFP–Swi6 construct showed an extended conformation (Fig. 3a). We reasoned that the proximity of the CFP tag to the CD perhaps disrupts the ARK loop–CD interaction. Consistent with this reasoning, CFP–Swi6 forms a ~ 10 -fold weaker dimer than Swi6^{WT} (Fig. 2b and Supplementary Fig. 6). To maintain the ARK loop–CD interaction, we moved the CFP tag to the carboxy terminus (Fig. 3b). Swi6–CFP has a similar dimerization constant to Swi6^{WT}, consistent with having an intact ARK loop–CD interaction (Supplementary Fig. 6), shows a more condensed structure compared to CFP–Swi6 and has a lower sedimentation coefficient (Fig. 3b and Supplementary Fig. 6). These results raised the possibility that the extended conformation of CFP–Swi6 reflects the open state (Fig. 2c), which is capable of binding methylated nucleosomes. We therefore visualized Swi6 bound to a methylated nucleosome using cryo-electron microscopy, and for comparison, visualized nucleosomes alone (Fig. 3c). On the basis of our previous biochemical knowledge, we applied two-fold symmetry to the Swi6–nucleosome complex to obtain the three-dimensional reconstruction (see also Supplementary Methods). For the nucleosome and the Swi6–nucleosome complexes, three-dimensional reconstructions were calculated using the nucleosome structure as an initial model to an overall resolution of ~ 15 Å and ~ 25 Å, respectively (Fig. 3c, Supplementary Fig. 7 and Methods).

The 25 Å resolution of the Swi6–nucleosome complex precludes conclusions about the detailed conformations of the bound Swi6 dimers. We instead further analysed the difference density between the complex and nucleosome (Fig. 3c). Although we cannot rule out that Swi6 binding alters nucleosome conformation, the difference density has roughly the mass (~ 125 kDa) of two Swi6 dimers (~ 150 kDa) as determined previously¹⁸. We thus assume that the difference density is mainly contributed by the bound Swi6 dimers. The putative location of the CD suggests that one CD engages an H3 tail and one CD protrudes out in solution (Fig. 3d). This arrangement of the CDs is compatible with the sticky ends architecture proposed previously¹⁸ (Fig. 3e). The putative location of the CSD dimer suggests that this domain may also engage the nucleosome (Fig. 3c, d). This possibility has also been previously suggested^{24–26}. To test it directly, we measured binding of the Swi6^{DimerX} to methylated nucleosomes, and observed that disruption of the CSD dimer decreases binding by tenfold (Fig. 4a and Supplementary Fig. 8).

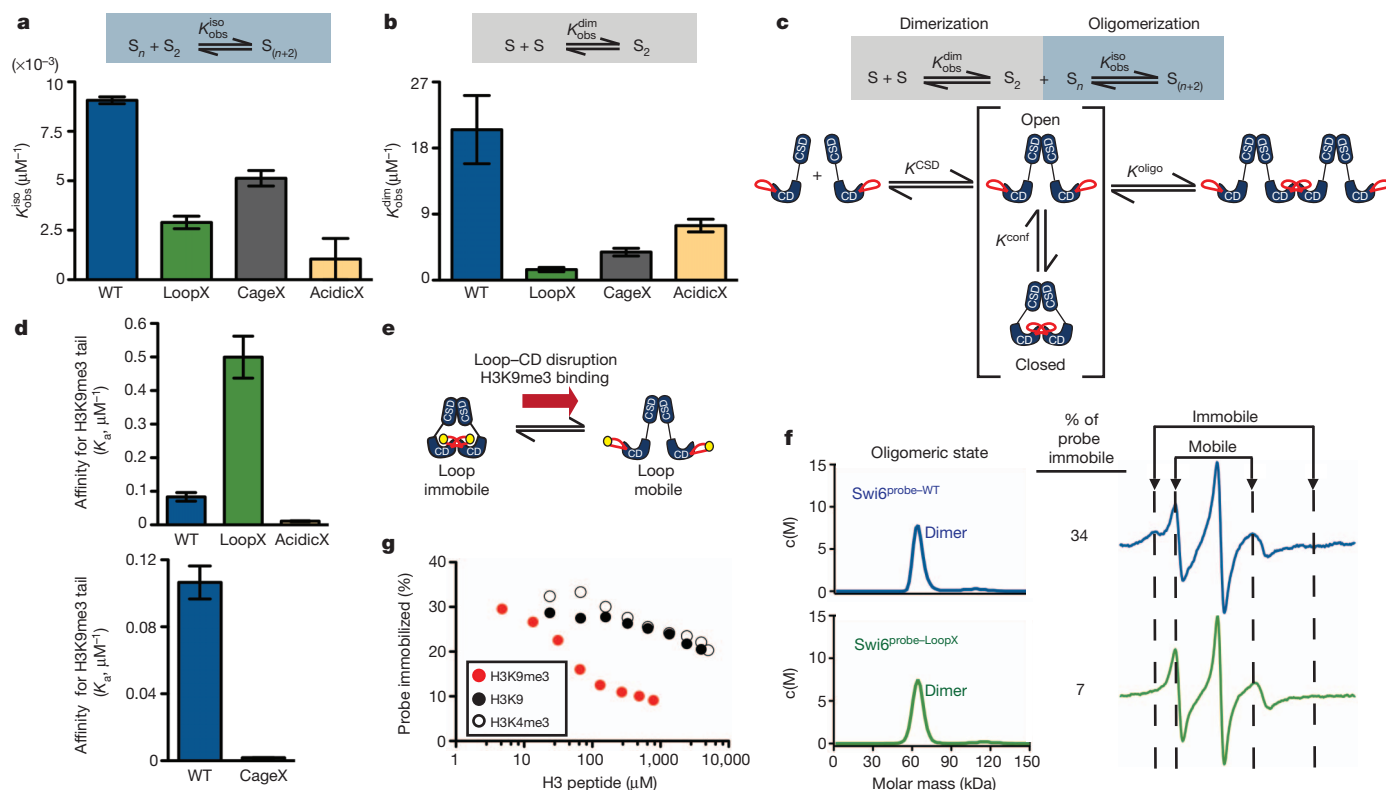


Figure 2 | Impact of disrupting H3 tail mimic-CD interaction.

a, b, Isodesmic association constant (K_{obs}^{iso}) (**a**) and dimerization association constant (K_{obs}^{dim}) (**b**) for Swi6 mutants (values in Supplementary Fig. 4). **c**, Model for the self-association of Swi6. ($K^{conf} = [open]/[closed]$). K^{oligo} is the isodesmic association constant for oligomerization from the open state. For Swi6^{WT}, $K_{obs}^{dim} = \frac{K^{CSD}}{K^{conf}}$ and $K_{obs}^{iso} = K^{conf} \times K^{oligo}$. **d**, Affinity constants (K_a) for H3K9me3 tail peptide measured by tryptophan fluorescence (top) and

fluorescence anisotropy (bottom) studies (values in Supplementary Fig. 8).

e, Location of MSL (4-maleimido-2,2,6,6-tetramethyl-1-piperidinyloxy) probe on G95C (yellow circle). **f**, Sedimentation velocity AUC (left panels) and EPR analyses (right panels) of Swi6^{probe-WT} and Swi6^{probe-LoopX}. Representative EPR spectra shown as derivative of absorbance (y axis) versus magnetic field (x axis). c(M), molar mass distribution. Errors for probe immobilized <10%.

g, Impact of 18-nucleotide H3 peptides on probe immobilization.

[Swi6^{probe-WT}] = 20 μM . For all panels, error bars represent s.e.m. ($n \geq 3$).

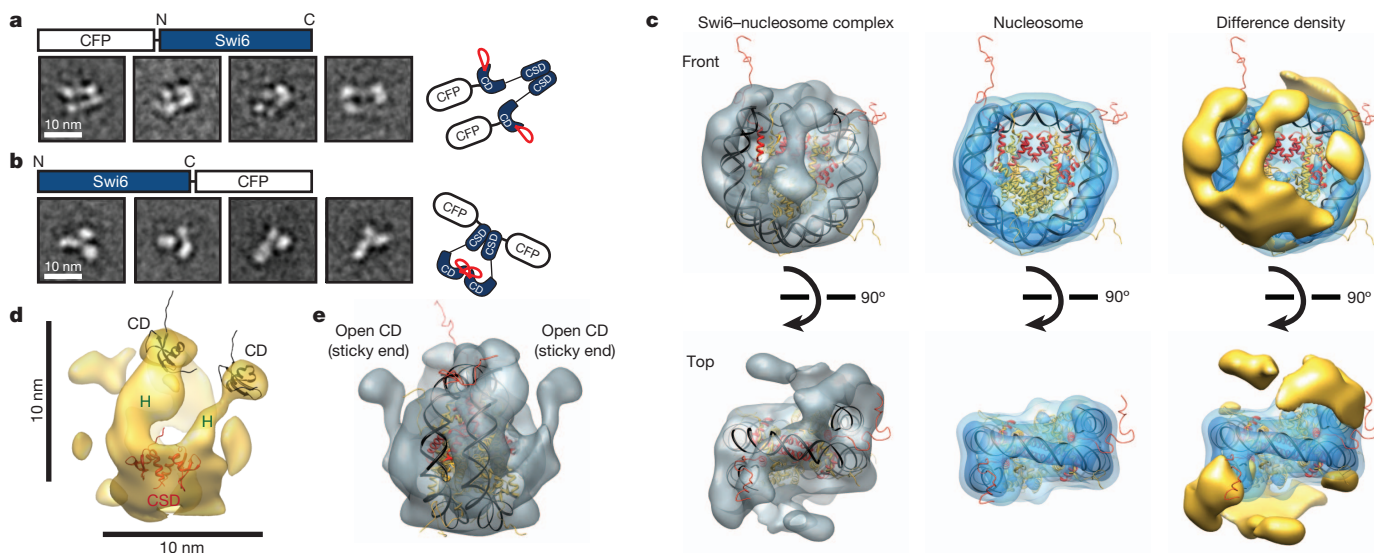


Figure 3 | Electron microscopy studies of Swi6 and Swi6-H3K9me3

nucleosome complex. **a, b**, CFP-Swi6 (**a**) and Swi6-CFP (**b**). Representative two-dimensional class averages are shown. **c**, Two different views of three-dimensional reconstruction of the Swi6-H3K9me3 (K_C indicates derivation of this analogue by alkylation of a cysteine residue, see Methods) nucleosome complex (left), nucleosome (middle) and difference map between the two reconstructions (right). Nucleosome crystal structure (PDB code, 1KX5) was

fitted into reconstruction. Iso-surface of nucleosome three-dimensional reconstruction at high threshold in dark blue and low threshold in light blue (nucleosome type used in Methods). H3, red; difference map, yellow.

d, Putative locations of Swi6 domains docked into difference map: CD of Swi6 (black; PDB code, 2RSO; amino acids 72–142), CSD domain of Swi6 (red; PDB code, 1E0B) and hinge (H). **e**, Proposed locations of the two unoccupied CDs.

Surprisingly, in contrast to the results with the H3 tail peptides (Fig. 2d), disrupting auto-inhibition through the $\text{Swi6}^{\text{LoopX}}$ mutant reduced binding to methylated nucleosomes by tenfold, even though discrimination for the methyl mark was maintained (Fig. 4a and Supplementary Fig. 8). This suggested that, when displaced from the CD, the ARK loop may help Swi6 make additional interactions with the nucleosome. We tested whether the positively charged ARK loop assists interactions with DNA. We found that Swi6^{WT} binds \sim fourfold tighter than $\text{Swi6}^{\text{LoopX}}$ to a 20-base pair DNA duplex (Fig. 4b). Further, Swi6^{WT} bound to the DNA \sim fourfold tighter with saturating H3K9me3 peptide, consistent with the loop being available when displaced from the CD (Fig. 4b and Supplementary Fig. 8). In contrast to $\text{Swi6}^{\text{LoopX}}$, $\text{Swi6}^{\text{AcidicX}}$, which binds H3K9me3 peptides more weakly than Swi6^{WT} , also binds methylated nucleosomes \sim sevenfold more weakly (data not shown).

On the basis of the above data, we propose that binding to methylated nucleosomes has two coupled effects: (1) release of ARK loops to directly or indirectly help DNA binding; and (2) release of two CDs that can bridge nearby nucleosomes (Fig. 4c). This new model revises our previously proposed model¹⁸ (see Supplementary Discussion). Our data imply that the cooperative action of the CD, the CSD–CSD dimer

and the ARK loop couples the assembly of Swi6 to the recognition of specific features of the nucleosomal template such as H3K9 methylation and nucleosomal DNA. This coupling can ensure correct targeting to H3K9-methylated chromatin and reduce aberrant spread in euchromatin. Interestingly, the loop that stabilizes the auto-inhibited state assists in binding nucleosomes when in the open state. These mutually exclusive roles of the loop may enable switch-like behaviour in HP1 spreading.

To test the relevance of this model *in vivo*, we investigated the impact of the LoopX and AcidicX mutants in assembling a functional heterochromatin structure. As these mutants concomitantly weaken oligomerization and nucleosome binding, we expected to observe loss-of-function effects *in vivo*. We first investigated effects on the silencing of a ura4^+ reporter gene inserted at the pericentromeric *imr* (innermost repeat) region (Fig. 4d). Both mutants show defects in silencing that are comparable to the swi6^+ deletion strain and that are not due to reductions in protein levels (Fig. 4d and Supplementary Fig. 9). Next, we investigated effects at endogenous centromeric *dg* repeats. In the absence of the RNA interference (RNAi) machinery, Swi6 is important for maintaining high levels of H3K9 methylation at *dg* repeats²⁷. Although deletion of RNAi components causes a small but reproducible decrease in H3K9 methylation, further deletion of *swi6* causes a much larger decrease in H3K9 methylation²⁷. We find that the *loopX* and *acidicX* mutants also show large decreases in H3K9 methylation in the absence of an RNAi component such as *dicer* (*dcr1*) (Fig. 4e and Supplementary Fig. 9). These data imply that the ARK loop–CD interaction is important for the integrity of H3K9-methylated heterochromatin *in vivo*. Our results with the LoopX and AcidicX mutants are also consistent with previous work showing that mutations in these regions affect mitotic stability and mating type switching¹⁷, both of which depend on the integrity of heterochromatin.

The ability of Swi6 to exist in more than one discrete conformational state may allow it to interact with different regulators through the CSD–CSD interface, the hinge or the ARK loop, and this could alter the stability and structure of the Swi6 –nucleosome platform (Fig. 4f). The Ala-Arg-Lys-Gly-Gly-Gly sequence is absent in the other *S. pombe* HP1 protein, Chp2, and this difference may in part explain the different biological roles of Chp2 and Swi6 (ref. 28). In Swi6 , the ARK loop stabilizes the auto-inhibited state even though the lysine is not methylated, presumably due to the high effective concentration of the ARK loop relative to its partner CD. However, human HP1 α contains just the Lys-Gly residues of the Ala-Arg-Lys-Gly-Gly-Gly sequence and, in this context, the lysine can be monomethylated *in vivo*²⁰. It is tempting to speculate that the methylation energetically compensates for the loss of the arginine while also making the interaction more regulatable. Protein assemblies that are controlled by release of auto-inhibition have been well characterized in processes such as actin nucleation and protein tyrosine kinase activation^{29,30}. We anticipate that similarly sophisticated mechanisms govern the assembly, spread and functions of HP1-mediated heterochromatin.

METHODS SUMMARY

Swi6 was purified from *Escherichia coli* as described previously¹⁸. Except where specified, all experiments were performed in 20 mM HEPES, pH 7.5, 150 mM KCl and 1 mM dithiothreitol. Nucleosomes were prepared using recombinant *Xenopus laevis* histones with and without a methyl lysine analogue at H3K9. Silencing assays were performed at 30 °C for 2–3 days. The Ab1220 (Abcam) was used for H3K9me2 chromatin immunoprecipitation.

Full Methods and any associated references are available in the online version of the paper.

Received 21 July 2012; accepted 22 February 2013.

Published online 13 March 2013.

1. Eissenberg, J. C. & Elgin, S. C. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **10**, 204–210 (2000).
2. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).

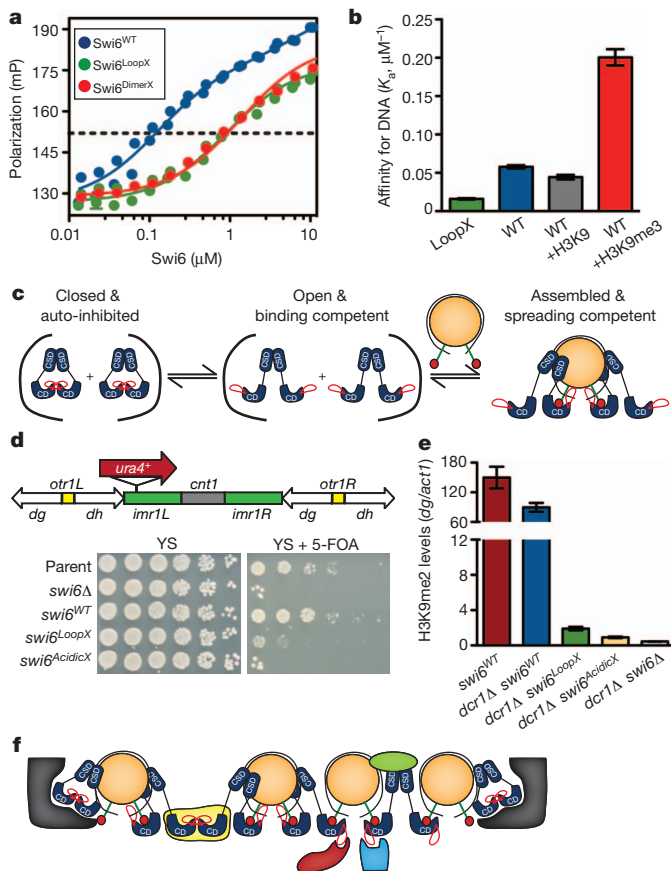


Figure 4 | Nucleosome recognition and *in vivo* impact of disrupting loop-CD interaction. **a**, **b**, Nucleosome binding assayed by fluorescence anisotropy (**a**) and affinity constants for 20-base-pair DNA (**b**). K_d values are in Supplementary Fig. 8. Error bars represent s.e.m. ($n \geq 3$). mP, millipolarization units. **c**, Model for conformational switch in Swi6 upon binding methylated nucleosomes. **d**, Top, schematics of centromere 1 showing ura4^+ reporter. cnt, centromere; L, left; otr, outer-most repeats; R, right. Bottom, silencing assay using ura4^+ reporter. **e**, $\text{swi6}^{\text{LoopX}}$ and $\text{swi6}^{\text{AcidicX}}$ mutants decrease H3K9 methylation levels at the centromeric *dg* on a *dcr1*Δ background. Error bars denote s.e.m. from three independent immunoprecipitations. **f**, Model depicts how conformational versatility of the HP1–chromatin platform enables recruitment of diverse regulators that promote (yellow, red, blue and green cartoons) or inhibit (grey cartoon) heterochromatin spread.

3. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
4. Noma, K., Allis, C. D. & Grewal, S. I. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**, 1150–1155 (2001).
5. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nature Rev. Genet.* **8**, 35–46 (2007).
6. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
7. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
8. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
9. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
10. Yamada, T., Fukuda, R., Himeno, M. & Sugimoto, K. Functional domain structure of human heterochromatin protein HP1(Hsa): involvement of internal DNA-binding and C-terminal self-association domains in the formation of discrete dots in interphase nuclei. *J. Biochem.* **125**, 832–837 (1999).
11. Brasher, S. V. *et al.* The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *EMBO J.* **19**, 1587–1597 (2000).
12. Cowieson, N. P., Partridge, J. F., Allshire, R. C. & McLaughlin, P. J. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Curr. Biol.* **10**, 517–525 (2000).
13. Smothers, J. F. & Henikoff, S. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr. Biol.* **10**, 27–30 (2000).
14. Mendez, D. L. *et al.* The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. *ChemBiochem* **12**, 1084–1096 (2011).
15. Zhao, T., Heyduk, T., Allis, C. D. & Eissenberg, J. C. Heterochromatin protein 1 binds to nucleosomes and DNA *in vitro*. *J. Biol. Chem.* **275**, 28332–28338 (2000).
16. Keller, C. *et al.* HP1^{Swi6} mediates the recognition and destruction of heterochromatic RNA transcripts. *Mol. Cell* (2012).
17. Wang, G. *et al.* Conservation of heterochromatin protein 1 function. *Mol. Cell. Biol.* **20**, 6970–6983 (2000).
18. Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell* **41**, 67–81 (2011).
19. Frigon, R. P. & Timasheff, S. Magnesium-induced self-association of calf brain tubulin. II. Thermodynamics. *Biochemistry* **14**, 4567–4573 (1975).
20. LeRoy, G. *et al.* Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. *Mol. Cell. Proteom.* **8**, 2432–2442 (2009).
21. Sampath, S. C. *et al.* Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol. Cell* **27**, 596–608 (2007).
22. Ruan, J. *et al.* Structural basis of the chromodomain of Cbx3 bound to methylated peptides from histone h1 and G9a. *PLoS ONE* **7**, e35376 (2012).
23. Rice, S. *et al.* A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
24. Dawson, M. A. *et al.* JAK2 phosphorylates histone H3Y41 and excludes HP1 α from chromatin. *Nature* **461**, 819–822 (2009).
25. Lavigne, M. *et al.* Interaction of HP1 and Brg1/Brm with the globular domain of histone H3 is required for HP1-mediated repression. *PLoS Genet.* **5**, e1000769 (2009).
26. Richart, A. N., Brunner, C. I. W., Stott, K., Murzina, N. V. & Thomas, J. O. Characterization of chromoshadow domain-mediated binding of heterochromatin protein 1 α (HP1 α) to histone H3. *J. Biol. Chem.* **287**, 18730–18737 (2012).
27. Sadaie, M., Iida, T., Urano, T. & Nakayama, J.-I. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *EMBO J.* **23**, 3825–3835 (2004).
28. Sadaie, M. *et al.* Balance between distinct HP1 family proteins controls heterochromatin assembly in fission yeast. *Mol. Cell. Biol.* **28**, 6973–6988 (2008).
29. Mullins, R. D. How WASP-family proteins and the Arp2/3 complex convert intracellular signals into cytoskeletal structures. *Curr. Opin. Cell Biol.* **12**, 91–96 (2000).
30. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Tretyakova for preparation of histone proteins and J. Leonard for sample preparation for cryo-electron microscopy of nucleosome alone. We thank W. Lim, M. Simon, K. Armache, J. Zalatan, L. Racki and members of the Narlikar laboratory for discussions. D.C. would like to thank I. Ortiz Torres and K. M. Kuchenbecker for scientific discussions and members of the Schuck laboratory for advice on AUC approaches. This work was supported by a grant from the Hillblom foundation to D.C., by grants from the American Cancer Society and Leukemia and Lymphoma Society to G.J.N., National Institutes of Health (NIH) grant R01GM071801 to H.D.M. and by a New Technology Award to Y.C. from the UCSF Program for Breakthrough Biomedical Research. P.S. was supported by the Intramural Research Program of the NIBIB, NIH. N.N. and E.P. were supported by the NIH grant AR053720.

Author Contributions D.C. and G.J.N. identified, developed and addressed the core questions. D.C. performed the bulk of the experiments. P.S. trained D.C. in the use of AUC approaches and was instrumental in interpreting the AUC data. N.N. performed the EPR experiments. D.B.M. trained D.C. in strain construction and in the use of *S. pombe* assays. J.F.G. constructed some of the *S. pombe* strains and performed initial *in vivo* experiments. E.P., R.C., A.L. and D.C. deconvolved the EPR spectra. S.W. generated the cryo-electron-microscopy reconstruction of the nucleosome alone. M.L. generated the EM reconstructions of the Swi6–nucleosome complex and the two-dimensional reconstructions of the CFP–Swi6 and Swi6–CFP constructs. M.L., S.W. and Y.C. analysed the electron microscopy data. Y.C. oversaw all of the electron microscopy studies. H.D.M. oversaw the design and interpretation of the *in vivo* experiments. R.C. oversaw the EPR analysis and interpretation. D.C. and G.J.N. wrote the bulk of the manuscript with substantial intellectual contributions from R.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.N. (geeta.narlikar@ucsf.edu).

METHODS

Protein cloning and purification. Swi6 proteins were purified from *Escherichia coli* as described previously¹⁸. Except for the CFP-tagged proteins, all other Swi6 protein purifications yield final proteins that are devoid of N- or C-terminal tags. Protein concentrations of all Swi6 construct samples were measured by ultraviolet absorption at 280 nm and calculated using the experimentally determined extinction coefficient (see AUC section). To ensure that there was minimal DNA contamination, we measured the 260/280 ratio for every purified protein, which on average was ~0.5.

Reaction buffer conditions. Except where specified, all experiments were performed in the reaction buffer consisting of 20 mM HEPES, pH 7.5, 150 mM KCl and 1 mM dithiothreitol.

Nucleosome assembly. Core nucleosomes were assembled on 147 base pairs (bp) of DNA using the 601 positioning sequence, containing a Pst1 site 18 bp in from the 5' end. For the cryo-electron microscopy of nucleosome-alone studies, 207 bp of DNA containing the 601 sequence at one end was used. All nucleosomes were prepared using recombinant *Xenopus laevis* histones and assembled as described previously³¹. Methyl lysine analogue containing H3 histones at position 9 (H3K9me3) were prepared as described previously³².

Tryptophan fluorescence studies. The association between Swi6 proteins and the H3 peptides (amino acids 1–18) were measured following the increases in the internal fluorescence of Trp 104 (one of the three residues in the aromatic cage) using an ISS K2 fluorimeter at 30 °C. Samples containing 200 nM Swi6 in reaction buffer were mixed with increasing concentrations of each H3 peptide, trimethylated or unmethylated at lysine 9. After an incubation for 10 min at 30 °C, the fluorescence of Trp 104 was measured with the incident wavelength of 295 nm. The observed fluorescence intensity (F_{obs}) at 330 nm was plotted as a function of peptide concentration. A 1:1 binding model was fit to the data using GraphPad Prism and the following set of equations:

$$F_{\text{obs}} = \frac{(F_{\text{max}}[\text{H3}_p] + F_{\text{min}}K_d)}{([\text{H3}_p] + K_d)}$$

in which F_{max} is the fluorescence at saturating peptide, F_{min} is the fluorescence in the absence of peptide and H3_p represents the H3 tail peptide. The obtained dissociation constant (K_d) values were averaged over three independent sets of data.

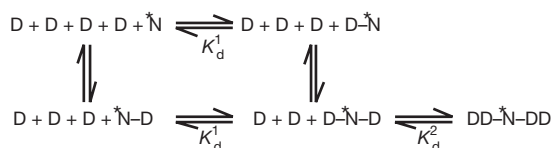
Fluorescence polarization studies. Fluorescence polarization-based measurements of binding to H3 tail peptides (amino acids 1–15), DNA and nucleosomes were performed in reaction buffer with 0.01% NP-40 at 24 °C. In total, 5–10 nM of peptide, DNA or nucleosomes were used and Swi6 concentrations were varied. The binding reaction was incubated for 30 min at 24 °C and fluorescence polarization was measured using an Analyst HT (Molecular Devices) with excitation and emission wavelengths of $\lambda_{\text{ex}} = 480$ nm and $\lambda_{\text{em}} = 530$ nm, respectively. The H3 peptide was labelled at the N terminus with a fluorescein probe (FAM). The peptide was synthesized by GenScript. The 20-base-pair DNA used in the DNA-binding assay was 5'-labelled with 5,6 carboxy-fluorescein (IDT). The DNA to assemble fluorescent nucleosomes was labelled on one end by amplifying the sequence using PCR with a primer covalently linked to 6-carboxyfluorescein by a 6-carbon linker (IDT). All data were analysed using GraphPad Prism.

The peptide and DNA-binding data were fit by the following equation:

$$F_{\text{obs}} = \frac{(F_{\text{max}}[\text{Swi6}] + F_{\text{min}}K_d)}{([\text{Swi6}] + K_d)}$$

In which F_{obs} is the fluorescence polarization signal observed, F_{min} is the fluorescence polarization signal for the probe alone (peptide or DNA) and F_{max} is the fluorescence polarization signal at saturating [Swi6]. The obtained K_d values were averaged over three or more independent sets of data.

The following model was used to fit the nucleosome-binding data to account for Swi6–Swi6 oligomerization that is scaffolded by the Swi6–nucleosome complex. Because the fluorescent probe is located only on one end of the DNA (asterisks), we made the assumption that changes in fluorescence polarization reflect binding of Swi6 on one side of the nucleosome. We proposed that binding to the other side occurs independently and is invisible to our assay:



In which D indicates a Swi6 dimer, D–N–D is the Swi6–nucleosome complex and DD–N–DD is the Swi6–nucleosome complex bound by additional Swi6

dimers. The fluorescence polarization nucleosome data were fitted using the following equation:

$$F_{\text{obs}} = \frac{(F_0 K_d^1 K_d^2 + F_1 [\text{Swi6}] K_d^2 + F_2 [\text{Swi6}]^2)}{(K_d^1 K_d^2 + [\text{Swi6}] K_d^2 + [\text{Swi6}]^2)}$$

In which F_{obs} is the fluorescence polarization signal observed, F_0 is the fluorescence polarization signal for the nucleosome alone, F_1 is the fluorescence polarization signal of the saturated Swi6–nucleosome complex, F_2 is the fluorescence polarization signal due to the oligomerization of Swi6 scaffolded by the Swi6–nucleosome complex, K_d^1 is the dissociation constant for the Swi6–nucleosome complex and K_d^2 is the dissociation constant for Swi6–Swi6 scaffolded by the Swi6–nucleosome complex. Data comparing Swi6^{WT} to the mutants were globally analysed: the F_0 , F_1 and F_2 were fixed among all proteins whereas the K_d^1 and K_d^2 were floated.

AUC studies. Swi6 proteins were individually dialysed into reaction buffer overnight. Swi6 proteins were quantified by ultraviolet absorption at 280 nm. We experimentally determined the Swi6^{WT} extinction coefficient by recording both interference fringes and ultraviolet absorbance at 280 nm for a given Swi6 sample. We then converted the number of interference fringes observed into milligrams of Swi6 per millilitre using an average refractive increment of 4.1 fringes mg^{-1} ml. Using this estimated concentration and the absorbance value at 280 nm, we then calculated the extinction coefficient at 280 nm to be 36,880 $\text{M}^{-1} \text{cm}^{-1}$. Simultaneous detection of protein by ultraviolet absorbance at multiple wavelengths allowed for the determination of the extinction coefficients at 230 nm and 250 nm (13,650 $\text{M}^{-1} \text{cm}^{-1}$ and 221,000 $\text{M}^{-1} \text{cm}^{-1}$, respectively).

All sedimentation experiments were conducted using an analytical ultracentrifuge (Beckman Coulter) equipped with either sole absorption optical scanner (Optima XLA) or both absorption and interference optics scanner (Optima XLI). Data were acquired with ProteomeLab data acquisition software 5. Global analysis of sedimentation equilibrium and sedimentation velocity isotherm data was performed using the SEDPHAT software. Error estimates were calculated on the basis of replicates of three or more experiments and confidence intervals based on F-statistics and the error projection method. Partial-specific volume (v), solution density (ρ) and solution viscosity (η) were calculated in SEDNTERP.

Sedimentation equilibrium. Sedimentation equilibrium experiments were conducted at 8 °C in an Optima XLI/A at rotor speeds of 6,000, 11,000 and 18,000 r.p.m. in double-sector centrepieces with sample volumes of 170 μl . Loading concentration of Swi6, in monomer units, was varied from 1.7 μM to 32 μM . Absorbance data, at wavelengths of 280, 250 and 230 nm, and immunofluorescence data were acquired from samples at five different loading concentrations at all rotor speeds. Global analysis of data at different wavelengths and rotor speeds was conducted with the software SEDPHAT, using Boltzmann exponentials representing the predicted concentration profiles of each species in chemical equilibrium, with amplitudes at all radii constrained by the mass action law:

$$c_{\text{tot}} = c_1 + n K c_1^n + \sum_{i=1}^{\infty} (n + im) K c_1^n (L c_1^i)^i$$

in combination with the method of implicit mass conservation, using the bottom position of each solution column as an adjustable parameter. In the above equation, c_{tot} is the total protein concentration, c_1 is the concentration of Swi6 monomer, n is the number of Swi6 subunits that self-associate in the first-step of association ($n = 2$ for dimer formation), m is the molecularity of the chain-elongation unit ($m = 2$ for a dimeric chain-elongation unit), K is the association constant for Swi6 dimerization ($K_{\text{obs}}^{\text{dim}}$), L is the association constant for Swi6 isodesmic chain elongation ($K_{\text{obs}}^{\text{iso}}$) and i is the number of units of m added to the initial complex formed by self-association of n Swi6 subunits³³. Summation of terms was carried out to a relative numerical precision of 10^{-6} .

Sedimentation velocity. Samples volumes of 400 μl at an overall final optical density (OD) between 0.1 and 1.0, were pipetted into double-sector centrepieces and inserted in an eight-hole rotor, which was placed in the temperature pre-equilibrated AUC chamber. An additional incubation period of 1–2 h was added with the rotor at rest and under vacuum for temperature equilibration. For experiments performed at 4 °C, the samples were left equilibrating under vacuum overnight. Runs were performed at a rotor speed of 50,000 r.p.m. for more than 12 h. Scans were collected following ultraviolet absorption at 230, 250 and 280 nm, scanned with a radial step size of 0.003 cm in continuous mode, and/or using the interference system. Data were analysed using a $c(s)$ continuous distribution of Lamm equation solutions with the software SEDFIT, followed by integration and assembly into an isotherm of weighted-average s -values. The isotherm was modelled in SEDPHAT with mass-action-based models for the weighted-average s -value

$$s_w(c_{\text{tot}}) = [c_1 s_1 + n K c_1^n s_n + \sum_i (n + im) K c_1^n (L c_1^m)^i s_n \left(\frac{n + im}{n} \right)^k] \frac{1}{c_{\text{tot}}} (1 - k_s c_{\text{tot}} m_1)$$

assuming a power law for the sedimentation coefficients of oligomeric species with $\kappa = 0.566$ (consistent with increasingly elongated oligomers; this value was pre-determined from the global fit of sedimentation equilibrium and sedimentation velocity on an extensive data set), in combination with an overall hydrodynamic non-ideality term of magnitude $k_s = 0.01 \text{ ml g}^{-1}$. As in the equation above, c_{tot} is the total protein concentration, c_1 is the concentration of Swi6 monomer, n is the number of Swi6 subunits that self-associate in the first-step of association ($n = 2$ for dimer formation), m is the molecularity of the chain elongation unit ($m = 2$ for a dimeric chain elongation unit), K is the association constant for Swi6 dimerization ($K_{\text{obs}}^{\text{dim}}$), L is the association constant for Swi6 isodesmic chain elongation ($K_{\text{obs}}^{\text{iso}}$), s_1 is the sedimentation coefficient of Swi6 monomer and s_n is the sedimentation coefficient of Swi6 dimer ($n = 2$).

Rationale for different temperatures. (1) To obtain a model for Swi6 self-association we performed sedimentation equilibrium and sedimentation velocity AUC studies at 8 °C. Sedimentation equilibrium experiments are ~1 week long, so a temperature of 8 °C was used to stabilize the protein. Global analysis of both sedimentation equilibrium and sedimentation velocity AUC at 8 °C allowed us to obtain a thermodynamic information for Swi6 self-association as well as hydrodynamic parameters for Swi6 monomer, dimer and oligomers that were used in all the sedimentation velocity experiments performed at higher temperatures. (2) To compare dimerization properties between Swi6 mutants, we had to perform the experiments at 30 °C because dimerization is too tight at lower temperatures. (3) To stabilize probe-labelled Swi6 proteins, the EPR and AUC experiments were done at 4 °C.

EPR studies. EPR measurements were performed with an EMX EPR spectrometer (Bruker Instruments). First derivative, X-band spectra were recorded in a high-sensitivity microwave cavity using 50-s, 100-G wide magnetic field sweeps. The instrument settings were as follows: microwave power, 25 mW; time constant, 164 ms; frequency, 9.83 GHz; modulation, 1 G at a frequency of 100 kHz. Each spectrum used in the data analysis was an average of 10–40 50-s sweeps from an individual experimental preparation. Swi6³⁵ was labelled by reacting the sole cysteine residue (either K94C or G95C) with the EPR probe 4-maleimido-2,2,6,6-tetramethyl-1-piperidinyloxy (MSL, Sigma Aldrich). The protein was first dialysed overnight in reaction buffer without dithiothreitol. It was then incubated with MSL using a twofold molar excess of MSL to protein concentration. The mixture was then left to react for 4 h at 4 °C. The excess label was removed by a microcon concentrator, followed by an additional overnight dialysis step into the above buffer. The protein sample was incorporated into a 25- μl capillary and the EPR spectrum was recorded. The temperature of the sample was controlled by blowing dry air (warm or cool) into the cavity and monitored using a thermistor placed close to the experimental sample. To stabilize the Swi6^{probe-WT} and Swi6^{probe} mutants we performed the EPR and AUC experiments at 4 °C.

The spectra were deconvoluted into mobile and immobile spectral components using the protocols described in ref. 34.

Electron microscopy and image processing. Negative-stain electron microscopy of CFP–Swi6 and Swi6–CFP. Proteins were dialysed overnight in reaction buffer. In total, 2.5 μl of CFP–Swi6 at 0.34 μM and of Swi6–CFP at 0.1 μM was absorbed to a glow-discharged copper grid coated with carbon film for 30 s followed by conventional negative stain with 0.75% uranyl formate. Images were collected using a Tecnai T12 microscope (FEI Company) with a LaB₆ filament and operated at 120-kV accelerating voltage. All images were recorded at a magnification of $\times 67,000$ with an UltraScan 4096 \times 4096 pixel CCD camera (Gatan).

All images were 2×2 pixel binned to the final pixel size of 3.46 Å before any further processing. A total of 5,000 and 3,000 particles for CFP–Swi6 and Swi6–CFP, respectively were selected from ~50 images using the display program SAMViewer (written by M. Liao). All subsequent image processing was performed using Spider³⁵ and FREALIGN.

Cryo-electron microscopy studies of the nucleosome and Swi6–nucleosome complex. Cryo-electron microscopy data were collected using Tecnai TF20 electron microscope equipped with a field emission gun (FEI Company) and operated at 120 kV (for the nucleosome) or at 200 kV (for the Swi6–nucleosome complex). Images were collected at a nominal magnification of $\times 62,000$ using a TemF816 8K \times 8K CMOS camera (TVIS).

Nucleosome alone. The nucleosome contained 60 bp of flanking DNA (147 bp of 601 sequence + 60 bp extra DNA) and did not contain the methyl lysine analogue on H3K9. All images were binned by a factor of 2 (2.39 Å/pixel) for further processing. Defocus values were determined for each micrograph using CTFFIND³⁷

and ranged from $-1.5 \mu\text{m}$ to $-3 \mu\text{m}$. A total of 13,629 particles were selected and classified into 100 two-dimensional class averages. Three-dimensional reconstructions were calculated and refined using GeFREALIGN³⁶. The initial model was generated by filtering the atomic structure of the nucleosome (PDB code, 1KX5) to 35 Å (command `pdb2mrc` from EMAN package)³⁸. The resolution was estimated to be ~16.5 Å, on the basis of Fourier shell correlation (FSC) = 0.5 criteria.

Swi6–H3K9me3 nucleosome complex. Swi6 was dialysed overnight in reaction buffer. The binding reaction was set such that, first, both nucleosome and Swi6 concentrations were above the K_d value measured by fluorescence polarization (FP) and second, the Swi6 concentration was sufficient to titrate all the nucleosomes as assayed by native gel shift. Those same conditions were used previously to measure the stoichiometry of the complex by sedimentation velocity AUC and are known to result in homogenous samples¹⁸.

A total of 5,000 particles were selected and classified into 200 two-dimensional class averages and all were included in the final three-dimensional reconstruction. The cryo-electron microscopy three-dimensional reconstruction of the nucleosome alone was low-pass filtered to 35 Å and used as the initial model for three-dimensional refinement of the complex. We used our previous biochemical knowledge to guide the structural analysis. We have previously shown that the complex of Swi6 with an H3K9 methylated nucleosome contains two Swi6 dimers¹⁸. Given the pseudo-two-fold symmetry in the positions of the H3 tails, the simplest model posits that the Swi6 dimers also bind in a pseudo-symmetric manner with one dimer on either side of the nucleosome. Indeed in some of the two-dimensional class averages we observe density on either side of the nucleosome consistent with the predictions of the biochemical analysis (Supplementary Fig. 7). We therefore applied two-fold symmetry to obtain the three-dimensional reconstruction. The resolution of the final three-dimensional reconstruction was estimated to be ~25 Å, on the basis of FSC = 0.5 criteria. This same resolution was also obtained when the cryo-electron microscopy three-dimensional reconstruction of nucleosome alone was low-pass filtered to 60 Å.

All three-dimensional reconstructions were visualized by UCSF Chimera. The 'Fit in Map' function of Chimera was used to dock the atomic structure of the nucleosome (PDB code, 1KX5) into the three-dimensional volume³⁹.

To calculate the difference map, the nucleosome alone and the Swi6–nucleosome complex maps were low-pass filtered to 25 Å. The difference map was calculated by subtracting the nucleosome from the Swi6–nucleosome complex using the program `diffmap.exe` (provided by N. Grigorieff), which normalizes the density maps before calculating the difference map. The extended shape of the difference density is compatible with the shape of the Swi6 dimer visualized in the negatively stained CFP–Swi6 dimer (Fig. 3a). Such similarity suggested a model for the arrangement of the individual domains of Swi6 and enabled us to manually place the known crystal structures of the CD and CSD into the difference density (Fig. 3d).

Silencing assays. The strains were grown overnight to saturation and diluted to OD_{600 nm} of 1 at the highest dilution. Serial dilutions were performed with dilution factor of 5 and cells were grown on non-selective (YS)- and 5-FOA (2 g l⁻¹ 5-fluoroorotic acid)-containing media for *ura4⁺* reporter at 30 °C for 2–3 days.

Quantifying Swi6 protein levels in vivo. Swi6 protein levels were quantified using polyclonal antibodies raised in rabbits by injecting recombinant Swi6.

Chromatin immunoprecipitation. The chromatin immunoprecipitation assay was performed as described previously⁴⁰. Cells were lysed at 4 °C by bead beating seven times for 1 min each with 2-min rests on ice. Chromatin fraction was sonicated 20 times for 30 s, each with a 1-min rest in between cycles using Bioruptor. Ab1220 (Abcam) was used for H3K9me2 chromatin immunoprecipitation and protein A Dynabeads were used in the washing steps.

- Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).
- Simon, M. D. Installation of site-specific methylation into histones using methyl lysine analogs. *Curr. Protocols Mol. Biol.* Ch 21:Unit 21.18.1–10 (2010).
- Vistica, J. *et al.* Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.* **326**, 234–256 (2004).
- Purcell, T. J. *et al.* Nucleotide pocket thermodynamics measured by EPR reveal how energy partitioning relates myosin speed to efficiency. *J. Mol. Biol.* **407**, 79–91 (2011).
- Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Li, X., Grigorieff, N. & Cheng, Y. GPU-enabled FREALIGN: accelerating single particle 3D reconstruction and refinement in Fourier space on graphics processors. *J. Struct. Biol.* **172**, 407–412 (2010).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

38. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
39. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
40. Rougemaille, M., Shankar, S., Braun, S., Rowley, M. & Madhani, H. D. Ers1, a rapidly diverging protein essential for RNA interference-dependent heterochromatic silencing in *Schizosaccharomyces pombe*. *J. Biol. Chem.* **283**, 25770–25773 (2008).

Structural basis of kynurenine 3-monooxygenase inhibition

Marta Amaral^{1,2,3,4}, Colin Levy¹, Derren J. Heyes¹, Pierre Lafite⁵, Tiago F. Outeiro^{3,4,6}, Flaviano Giorgini², David Leys¹ & Nigel S. Scrutton¹

Inhibition of kynurenine 3-monooxygenase (KMO), an enzyme in the eukaryotic tryptophan catabolic pathway (that is, kynurenine pathway), leads to amelioration of Huntington's-disease-relevant phenotypes in yeast, fruitfly and mouse models^{1–5}, as well as in a mouse model of Alzheimer's disease³. KMO is a flavin adenine dinucleotide (FAD)-dependent monooxygenase and is located in the outer mitochondrial membrane where it converts L-kynurenine to 3-hydroxykynurenine. Perturbations in the levels of kynurenine pathway metabolites have been linked to the pathogenesis of a spectrum of brain disorders⁶, as well as cancer^{7,8} and several peripheral inflammatory conditions⁹. Despite the importance of KMO as a target for neurodegenerative disease, the molecular basis of KMO inhibition by available lead compounds has remained unknown. Here we report the first crystal structure of *Saccharomyces cerevisiae* KMO, in the free form and in complex with the tight-binding inhibitor UPF 648. UPF 648 binds close to the FAD cofactor and perturbs the local active-site structure, preventing productive binding of the substrate L-kynurenine. Functional assays and targeted mutagenesis reveal that the active-site architecture and UPF 648 binding are essentially identical in human KMO, validating the yeast KMO–UPF 648 structure as a template for structure-based drug design. This will inform the search for new KMO inhibitors that are able to cross the blood–brain barrier in targeted therapies against neurodegenerative diseases such as Huntington's, Alzheimer's and Parkinson's diseases.

There is great interest in the causative role of kynurenine pathway metabolites in neurodegenerative disorders such as Huntington's and Alzheimer's diseases⁶. Several of these metabolites are neuroactive: quinolinic acid (QUIN) is an excitotoxin^{10,11}, 3-hydroxykynurenine (3-HK) generates free radicals¹², xanthurenic and cinnabarinic acids activate metabotropic glutamate receptors^{13,14} and kynurenic acid (KYNA) is a neuroprotectant⁶. KMO lies at a critical branching point in the pathway between the synthesis of 3-HK/QUIN and KYNA (Fig. 1a) and its activity has a role in the neurotoxic and neuroprotective potential of the pathway. In the brain, KMO is expressed at low levels in neurons¹⁵ and is predominantly expressed in microglia^{1,16}, the resident immune cells of the CNS, suggesting a link between KMO function and inflammatory processes in the brain.

Inhibition of KMO activity leads to amelioration of several disease-relevant phenotypes in yeast, fruitfly and mouse models^{1–5}. Increased levels of KYNA relative to neurotoxic metabolites seem critical for this protection. Restoring endogenous levels of 3-HK to fruitflies lacking KMO activity eliminates this neuroprotection⁴, highlighting beneficial effects of 3-HK reduction due to KMO inhibition. In addition, pharmacological inhibition of KMO is neuroprotective in animal models of cerebral ischaemia^{17,18}, reduces dystonia in a genetic model of paroxysmal dyskinesia¹⁹, improves levodopa-induced dyskinesia in parkinsonian monkeys²⁰ and extends lifespan in a mouse model of cerebral

malaria²¹. Therefore, inhibition of KMO activity is an attractive therapeutic strategy for several acute and chronic neurological diseases⁶.

Despite interest in targeting KMO only a few potent inhibitors are available, and none appreciably penetrate the blood–brain barrier in

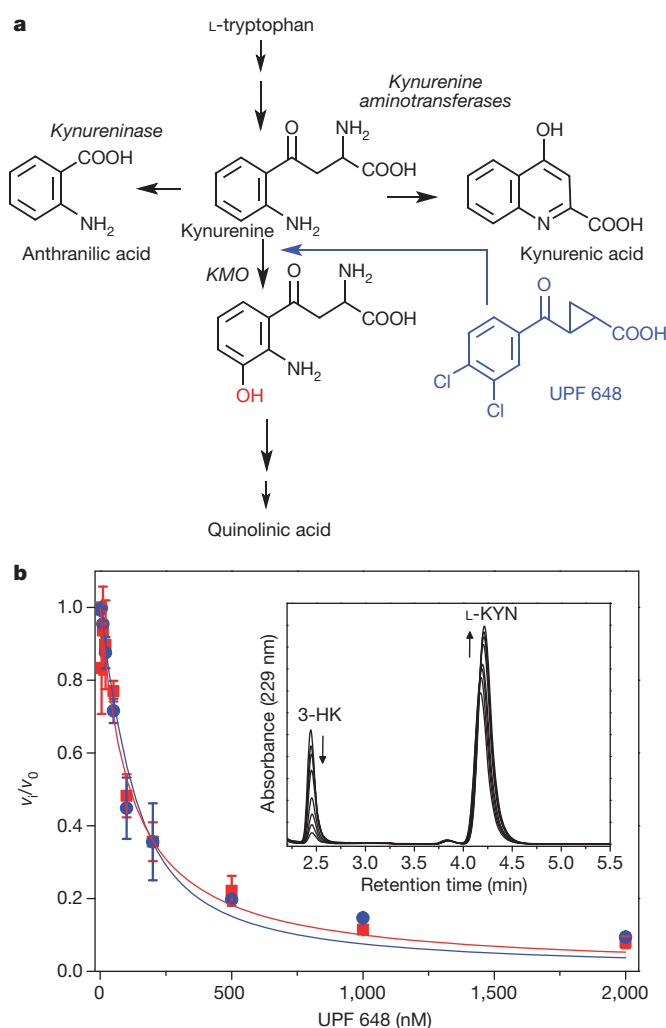


Figure 1 | Inhibition of KMO by UPF 648. **a**, Schematic overview of L-KYN metabolism. The KMO inhibitor UPF 648 is shown in blue. The hydroxyl moiety introduced by KMO is shown in red. **b**, Fractional velocity of 3-HK formation as a function of UPF 648 concentration with human and *S. cerevisiae* KMO (blue circles, human KMO; red squares, *S. cerevisiae* KMO). Error bars are standard deviation of three replica points. Inset, HPLC elution curves of product (3-HK) and substrate (L-KYN) at varied UPF 648 concentrations.

¹Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. ²Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. ³Cell and Molecular Neuroscience Unit, Instituto de Medicina Molecular, 1659-028 Lisboa, Portugal. ⁴Instituto de Fisiologia, Faculdade de Medicina da Universidade de Lisboa, 1649-028 Lisboa, Portugal. ⁵Institut de Chimie Organique et Analytique, Université d'Orléans, CNRS UMR 7311 BP6769, Rue de Chartres, 45067 Orléans Cedex 2, France. ⁶Department of Neurodegeneration and Restorative Research, Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center Göttingen, Waldweg 33, 37073 Göttingen, Germany.

Table 1 | Kinetic and dissociation constants for human, *S. cerevisiae* and active-site variants of KMO

Enzyme	K_i app UPF 648 (nM)	K_d UPF 648 (μ M)
hKMO*	56.7 \pm 8	—
ScKMO(Δ 394)	74 \pm 14	0.14 \pm 0.01
ScKMO(Δ 394, R83A) [†]	—	3.1 \pm 0.2
ScKMO(Δ 394, R83M) [†]	—	3.2 \pm 0.2

* Dissociation constant (K_d) for the human (h)KMO–UPF 648 complex could not be determined by fluorescence emission owing to low KMO expression yield.

[†] Kinetic assays to calculate the apparent inhibition constant (K_i app) of *S. cerevisiae* (Sc)KMO variants by UPF 648 could not be carried out because enzyme activity was substantially compromised after mutagenesis. As an alternative, fluorescence-emission measurements were performed to determine K_d by ligand perturbation of flavin fluorescence.

adult animals^{3,22}. One of these, UPF 648, has a half-maximum inhibitory concentration (IC_{50}) of 20 nM and provides protection against intrastriatal QUIN injections in kynurenine aminotransferase (KATII, also known as AADAT)-deficient mice²³. UPF 648 treatment also shifts kynurenine pathway metabolism towards enhanced neuroprotective KYNA formation^{4,24}, and ameliorates disease-relevant phenotypes in a fruitfly model of Huntington's disease⁴. That known inhibitors do not cross the blood–brain barrier is an impediment to KMO-targeted drug discovery. KMO structures in complex with tight-binding inhibitors are required to design small-molecule inhibitors that can penetrate the blood–brain barrier. With this in mind, we determined the crystal structure of yeast KMO complexed with UPF 648. This enzyme–inhibitor structure can now be used to help develop new inhibitors of highly related human KMO.

We expressed full-length human KMO using the insect cell baculovirus system, which yielded small quantities (0.5 mg per litre of culture) of detergent-solubilized active KMO. The recombinant form had similar kinetic constants to native KMO from pig liver mitochondria²⁵. UPF 648 binds tightly to recombinant KMO (inhibition constant (K_i) = 56.7 nM). Poor stability and low expression yields, however, prevented crystallization. We therefore turned to *S. cerevisiae* KMO, which is related to human KMO (38% identity and 51% similarity). Expression of full-length *S. cerevisiae* KMO yielded a protein fragment (KMO(Δ 396)) with a lower molecular weight than anticipated. Electrospray ionization mass spectrometry indicated proteolytic cleavage at residue 396. Subsequently, we isolated a KMO(Δ 394) (residues 394–460 deleted) version of the enzyme engineered by site-directed mutagenesis (Methods) to define the cleavage point before crystallization (Supplementary Fig. 1 and Supplementary Table 1). The KMO(Δ 394) enzyme was active (Supplementary Figs 2 and 3), generated authentic

3-HK in high-performance liquid chromatography (HPLC)-based assays (Fig. 1b) and was inhibited by UPF 648 (K_i = 74 nM) with a potency similar to that seen with human KMO (Fig. 1b and Table 1).

The structure of the proteolysed form of yeast KMO (KMO(Δ 396)) was determined using selenomethionine single anomalous diffraction (Protein Data Bank (PDB) codes 4J2W and 4J31). We also solved structures of KMO(Δ 394) to 1.85 Å resolution. The final model contains residues 1–97 and 101–390 and the bound FAD cofactor. Both crystal forms contain a putative KMO dimer in the asymmetric unit (Fig. 2a). The KMO fold is similar to other flavin-dependent hydroxylase structures^{26,27}, with highest structural similarity to 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase²⁸ (root mean squared deviation, 2.3 Å over 310 C α atoms; overall sequence identity, 16%; Q-score, 0.43; Z-score, 15.0). An overlay of individual KMO monomers reveals variation in the position of the carboxy-terminal α -helix, with most monomers showing disorder beyond residue 380. The linker region following the second strand of the antiparallel β -sheet involved in substrate binding is disordered, with large variations in the positions of residues 96–97 and 101–104. The relative position of the FAD-binding domain and six-stranded antiparallel β -sheet domain is subject to minor variation, reminiscent of domain motion coupled to substrate binding in other members of this family²⁶.

In absence of substrate, flexibility in relative positioning of both domains flanking the KMO active site is also reflected in distinct conformations observed for residues lining the active site (Fig. 3). The *re*-face of the FAD is connected to solvent by a narrow water-filled cavity that runs perpendicular to the active-site cleft (Supplementary Fig. 4). A structural water is located above the FAD C4a, mimicking the position of the C4a–peroxide intermediate formed upon reaction with oxygen. The dimethylbenzene moiety of the FAD isoalloxazine is protected from solvent by Lys 48 and the conserved residue Tyr 195 (Supplementary Fig. 4). In absence of large protein rearrangements, this suggests that a 'waving flavin' motion as demonstrated in other FAD-dependent monooxygenases²⁹ is unlikely to occur in KMO during turnover.

We were unable to obtain a KMO complex with L-kynurenine (L-KYN) but succeeded in co-crystallizing with UPF 648 (Supplementary Table 3, PDB code 4J36). The asymmetric unit contains a putative KMO dimer with one monomer containing UPF 648 bound in the active site, adjacent to the FAD *re*-face (Figs 2b and 3). The UPF 648 carboxylate is bound by conserved residues Arg 83 and Tyr 97 whereas the aromatic dichlorobenzene moiety is flanked by several hydrophobic residues

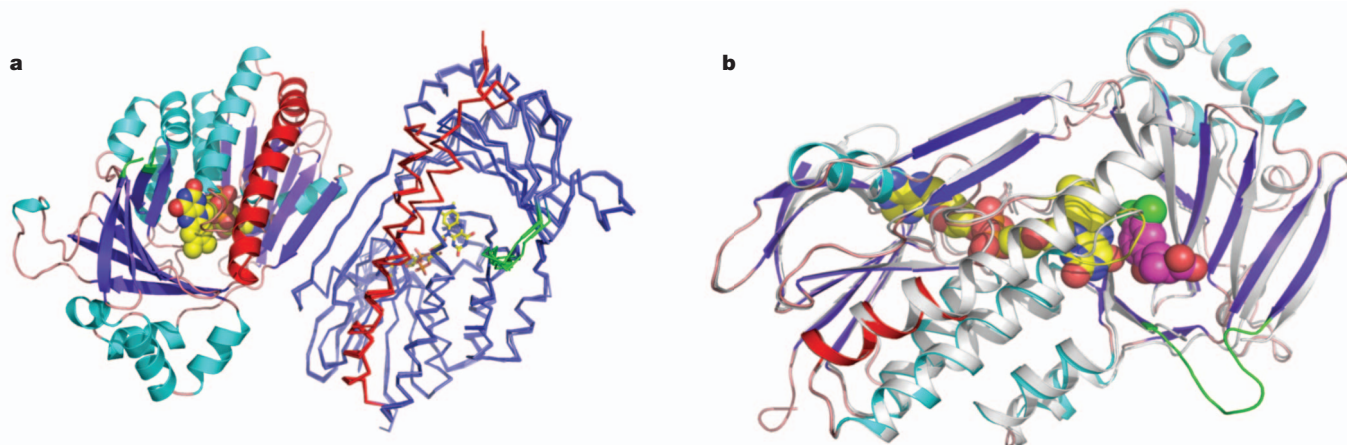


Figure 2 | Yeast KMO crystal structure. **a**, Depiction of the *S. cerevisiae* KMO dimer. One monomer is represented in cartoon format coloured according to secondary structure and the second monomer is displayed in ribbon format, with an overlay of the five ligand-free KMO monomers. For the second monomer, the C-terminal region (residue 348 onwards) is coloured red and the flexible linker region (residues 96–104) is depicted in green. **b**, Depiction of

changes induced by binding of the KMO inhibitor UPF 648. An overlay of both monomers from the KMO–UPF 648 structure is presented (PDB code 4J36), the unbound monomer coloured in grey. The UPF-bound monomer is coloured as in **a**. UPF 648 and FAD are shown as spheres colour-coded according to atom type.

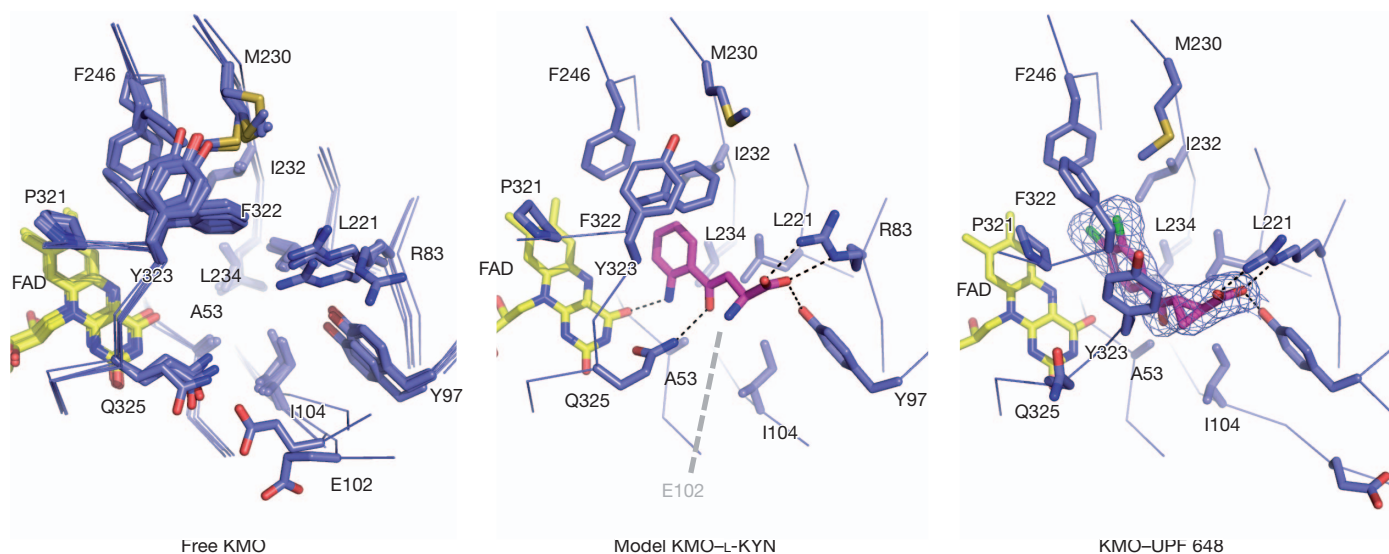


Figure 3 | The *Saccharomyces cerevisiae* KMO active site. Left panel shows an overlay of free enzyme structures obtained from various crystal forms, right panel depicts the KMO–UPF 648 interaction and middle panel shows a model

of the KMO–L-KYN complex. Electron density is shown for UPF 648 ($2F_o - F_c$ contoured at 1σ).

(Leu 221, Met 230, Ile 232, Leu 234, Phe 246, Pro 321, Phe 322), which are conserved in many KMO enzymes. UPF 648 binding induced structural changes in the enzyme, notably reorientation of the Pro 321–Gln 325 loop flanking the *re*-side of the FAD. A minor reorientation in the position of the six-stranded antiparallel β -sheet domain with respect to the flavin-binding domain is also evident. These changes result in increased disorder of the C-terminal α -helix, which is only visible up to Arg 359 (Fig. 2b). Reorientation of the Pro 321–Gln 325

loop is a consequence of the active site adapting to the presence of vicinal chloride substituents in UPF 648, neither of which have a counterpart in L-KYN. To provide sufficient space for both chlorides, Phe 322 moves away from the active site, effectively occupying a position previously taken by Tyr 233. The Pro 321–Gln 325 loop reorients to compensate for the altered Phe 322 position. This loop lines the postulated oxygen-binding site above the *re*-side of the FAD, which is effectively destroyed on binding UPF 648. Binding UPF 648 was found to accelerate hydrogen peroxide formation by a factor of ~ 20 -fold compared to reactions in absence of UPF 648 (Methods). This indicates a destabilization of the flavin C4a–hydroperoxide intermediate formed in the natural catalytic cycle of flavin monooxygenases in the presence of UPF 648 (Supplementary Table 2).

The chemical similarity of UPF 648 and L-KYN allowed modelling of L-KYN in the KMO active site (Methods). Modelling suggests that L-KYN is bound similarly, but without effect on the Pro 321–Gln 325 loop. The aromatic substrate moiety is located in the conserved hydrophobic pocket (residues Leu 221, Met 230, Ile 232, Leu 234, Phe 246, Pro 321, Phe 322) on the *re*-face of the flavin (Fig. 3). Additional polar contacts are formed between the conserved Gln 325 and the L-KYN carbonyl group, and between the substrate aniline nitrogen atom and the FAD O4 atom. Whereas the amino acid carboxylate is bound by Arg 83 and Tyr 97, the amino group is devoid of direct interactions with protein in the model. An additional salt bridge may be made between the L-KYN amine and the side chain of Glu 102, which is located in a highly flexible region of KMO. However, this residue is often replaced by a glutamine in other KMO enzymes (Supplementary

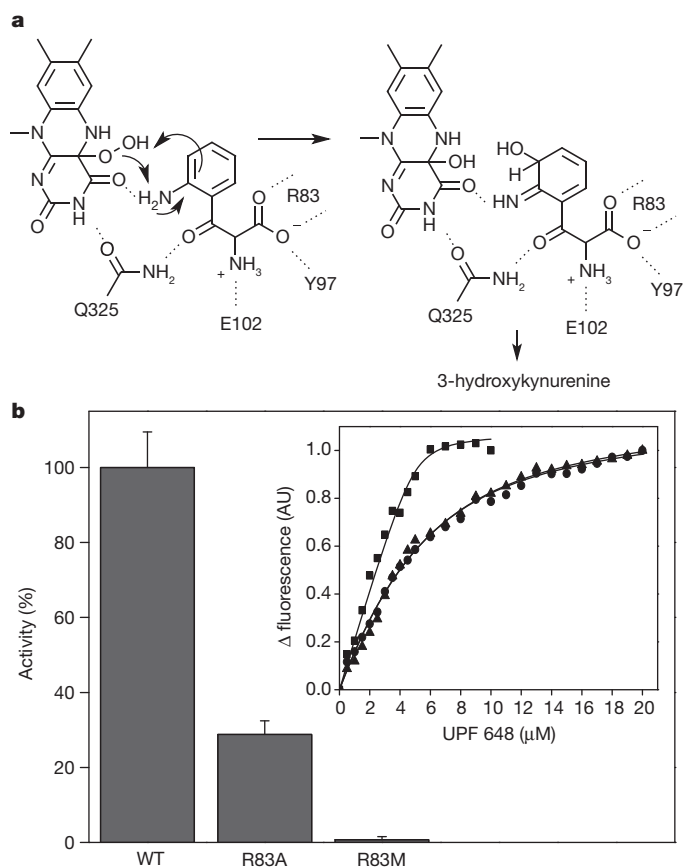


Figure 4 | Mechanism and importance of Arg 83. **a**, The proposed KMO L-KYN-hydroxylation mechanism based on the *S. cerevisiae* KMO–L-KYN model. After hydroxylation, the intermediate rearranges to form the 3-HK product. **b**, Enzymatic activity comparison between wild-type KMO and Arg 83 mutants. Enzyme activity is significantly reduced following mutation ($<29\%$ and $<1\%$ of wild-type activity for Ala 83 and Met 83 mutant KMO enzymes, respectively). Inset, measurement of binding constant for KMO inhibitor UPF 648 by ligand perturbation of flavin fluorescence emission. Titration of UPF 648 to $5\ \mu\text{M}$ wild-type KMO (closed squares), R83M (closed triangles) and R83A (closed circles) mutants resulted in perturbation of $520\ \text{nm}$ fluorescence emission. Fluorescence changes at $520\ \text{nm}$ as function of UPF 648 concentration fitted to the Morrison equation yields an observed $K_d = 137.8 \pm 8\ \text{nM}$ for wild type, $K_d = 3.1 \pm 0.2\ \mu\text{M}$ for R83M and $K_d = 3.2 \pm 0.2\ \mu\text{M}$ for R83A enzymes. Error bars represent the standard deviation of three replica points. AU, arbitrary units.

Fig. 5), suggesting that this interaction is not critical for enzyme activity. The model places the substrate C3 atom adjacent to the flavin C4a, where it is poised to attack the flavin C4a–peroxide intermediate (Fig. 4a).

All residues implicated by the KMO–L-KYN model as being involved in L-KYN binding are conserved across KMOs (Supplementary Fig. 5). We validated this model by mutating residue Arg83 (replaced by Ala83 and Met83) and performing inhibitor binding/kinetic assays (Fig. 4b). Enzyme activity is compromised following mutation (25% and <3% of wild-type activity for Ala83 and Met83 mutant KMO enzymes, respectively) as predicted by the KMO–L-KYN model. Mutation led to ~20-fold increase in dissociation constant (K_d) for the KMO–UPF 648 inhibitor complex, which implicates Arg83 in inhibitor binding.

Elucidation of the KMO crystal structure in free form and in complex with an established inhibitor is a major breakthrough for new KMO inhibitor design. This will permit docking screens using virtual compound libraries that may ultimately identify novel inhibitor scaffolds. Such studies will enable the design of new inhibitors that possess the selectivity and affinity to open up new opportunities for therapeutic intervention. Critically, this should inform the development of brain-penetrant KMO inhibitors. The KMO inhibitor Ro 61-8048 and prodrug JM6 have shown preclinical promise in animal models of neurodegeneration^{2–4,16–20}, but they do not appreciably penetrate the blood–brain barrier. These compounds probably confer neuroprotection by inhibiting KMO in the blood and consequent active transport of L-KYN into the brain, ultimately yielding increased levels of the neuroprotective metabolite KYNA^{3,30}. As this neuroprotection seems to be independent of KMO inhibition in the brain and reduction of downstream neurotoxic metabolite levels³, delivery of KMO inhibitors directly to the CNS should provide critical additional therapeutic efficacy and broaden the scope of disorders amenable to targeting.

METHODS SUMMARY

Discrimination assays. Human and *S. cerevisiae* KMO were produced using baculovirus and *Escherichia coli* expression systems, respectively. HPLC was used to monitor inhibition of KMO by UPF 648 by monitoring formation of product 3-HK across a range of inhibitor UPF 648 concentrations (0–2 μ M). Flavin fluorescence-emission measurements were used to determine K_d for KMO ligands through ligand perturbation of the fluorescence of enzyme-bound FAD. Detailed protocols can be found in Methods.

Crystallization and diffraction data collection. Crystals of *S. cerevisiae* KMO were obtained using the sitting-drop vapour-diffusion technique. Details describing the crystallogensis can be found in Methods. Diffraction data from *S. cerevisiae* KMO crystals were collected from single cryo-frozen samples at the Diamond Light Source, Harwell, UK. A single-wavelength anomalous dispersion data set at the selenium edge was used for initial phase determination. Data collection and refinement statistics are presented in Supplementary Table 3.

Full Methods and any associated references are available in the online version of the paper.

Received 18 December 2012; accepted 22 February 2013.

Published online 10 April 2013.

- Giorgini, F. *et al.* Histone deacetylase inhibition modulates kynurenine pathway activation in yeast, microglia, and mice expressing a mutant huntingtin fragment. *J. Biol. Chem.* **283**, 7390–7400 (2008).
- Giorgini, F., Guidetti, P., Nguyen, Q., Bennett, S. C. & Muchowski, P. J. A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. *Nature Genet.* **37**, 526–531 (2005).
- Zwilling, D. *et al.* Kynurenine 3-monooxygenase inhibition in blood ameliorates neurodegeneration. *Cell* **145**, 863–874 (2011).
- Campesan, S. *et al.* The kynurenine pathway modulates neurodegeneration in a *Drosophila* model of Huntington's disease. *Curr. Biol.* **21**, 961–966 (2011).
- Green, E. W. *et al.* *Drosophila* eye color mutants as therapeutic tools for Huntington disease. *Fly* **6**, 117–120 (2012).
- Schwarcz, R., Bruno, J. P., Muchowski, P. J. & Wu, H.-Q. Kynurenines in the mammalian brain: when physiology meets pathology. *Nature Rev. Neurosci.* **13**, 465–477 (2012).
- Platten, M., Litzénburger, U. & Wick, W. The aryl hydrocarbon receptor in tumor immunity. *Oncotarget* **1**, 396–397 (2012).

- Liu, X., Newton, R. C., Friedman, S. M. & Scherle, P. A. Indoleamine 2,3-dioxygenase, an emerging target for anti-cancer therapy. *Curr. Cancer Drug Targets* **9**, 938–952 (2009).
- Filippini, P. *et al.* Emerging concepts on inhibitors of indoleamine 2,3-dioxygenase in rheumatic diseases. *Curr. Med. Chem.* **19**, 5381–5393 (2012).
- Stone, T. W. & Perkins, M. N. Quinolinic acid: a potent endogenous excitant at amino acid receptors in CNS. *Eur. J. Pharmacol.* **72**, 411–412 (1981).
- Schwarcz, R., Whetsell, W. O. Jr & Mangano, R. M. Quinolinic acid: an endogenous metabolite that produces axon-sparing lesions in rat brain. *Science* **219**, 316–318 (1983).
- Okuda, S., Nishiyama, N., Saito, H. & Katsuki, H. Hydrogen peroxide-mediated neuronal cell death induced by an endogenous neurotoxin, 3-hydroxykynurenine. *Proc. Natl Acad. Sci. USA* **93**, 12553–12558 (1996).
- Copeland, C. S., Neale, S. A. & Salt, T. E. Actions of xanthurenic acid, a putative endogenous group II metabotropic glutamate receptor agonist, on sensory transmission in the thalamus. *Neuropharmacology* **66**, 133–142 (2013).
- Fazio, F. *et al.* Cinnabarinic acid, an endogenous metabolite of the kynurenine pathway, activates type 4 metabotropic glutamate receptors. *Mol. Pharmacol.* **81**, 643–656 (2012).
- Guillemin, G. J. *et al.* Characterization of the kynurenine pathway in human neurons. *J. Neurosci.* **27**, 12884–12892 (2007).
- Guillemin, G. J., Smith, D. G., Smythe, G. A., Armati, P. J. & Brew, B. J. Expression of the kynurenine pathway enzymes in human microglia and macrophages. *Adv. Exp. Med. Biol.* **527**, 105–112 (2003).
- Cozzi, A., Carpenedo, R. & Moroni, F. Kynurenine hydroxylase inhibitors reduce ischemic brain damage: studies with (*m*-nitrobenzoyl)-alanine (mNBA) and 3,4-dimethoxy-[*N*-(4-nitrophenyl)thiazol-2-yl]-benzenesulfonamide (Ro 61-8048) in models of focal or global brain ischemia. *J. Cereb. Blood Flow Metab.* **19**, 771–777 (1999).
- Moroni, F. *et al.* Studies on the neuroprotective action of kynurenine monooxygenase inhibitors in post-ischemic brain damage. *Adv. Exp. Med. Biol.* **527**, 127–136 (2003).
- Richter, A. & Hamann, M. The kynurenine 3-hydroxylase inhibitor Ro 61-8048 improves dystonia in a genetic model of paroxysmal dyskinesia. *Eur. J. Pharmacol.* **478**, 47–52 (2003).
- Samadi, P. *et al.* Effect of kynurenine 3-hydroxylase inhibition on the dyskinetic and antiparkinsonian responses to levodopa in Parkinsonian monkeys. *Mov. Disord.* **20**, 792–802 (2005).
- Clark, C. J. *et al.* Prolonged survival of a murine model of cerebral malaria by kynurenine pathway inhibition. *Infect. Immun.* **73**, 5249–5251 (2005).
- Reinhart, P. H. & Kelly, J. W. Treating the periphery to ameliorate neurodegenerative diseases. *Cell* **145**, 813–814 (2011).
- Sapko, M. T. *et al.* Endogenous kynurenate controls the vulnerability of striatal neurons to quinolinate: implications for Huntington's disease. *Exp. Neurol.* **197**, 31–40 (2006).
- Ceresoli-Borroni, G., Guidetti, P., Amori, L., Pellicciari, R. & Schwarcz, R. Perinatal kynurenine 3-hydroxylase inhibition in rodents: pathophysiological implications. *J. Neurosci. Res.* **85**, 845–854 (2007).
- Uemura, T. & Hirai, K. L-kynurenine 3-monooxygenase from mitochondrial outer membrane of pig liver: purification, some properties, and monoclonal antibodies directed to the enzyme. *J. Biochem.* **123**, 253–262 (1998).
- Palfey, B. A. & McDonald, C. A. Control of catalysis in flavin-dependent monooxygenases. *Arch. Biochem. Biophys.* **493**, 26–36 (2010).
- van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* **124**, 670–689 (2006).
- McCulloch, K. M., Mukherjee, T., Begley, T. P. & Ealick, S. E. Structure of the PLP degradative enzyme 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase from *Mesorhizobium loti* MAFF303099 and its mechanistic implications. *Biochemistry* **48**, 4139–4149 (2009).
- Ortiz-Maldonado, M., Ballou, D. P. & Massey, V. A rate-limiting conformational change of the flavin in *p*-hydroxybenzoate hydroxylase is necessary for ligand exchange and catalysis: studies with 8-mercapto- and 8-hydroxy-flavins. *Biochemistry* **40**, 1091–1101 (2001).
- Fukui, S., Schwarcz, R., Rapoport, S. I., Takada, Y. & Smith, Q. R. Blood–brain barrier transport of kynurenines: implications for brain synthesis and metabolism. *J. Neurochem.* **56**, 2007–2017 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Schwarcz for supplying UPF 648. We also thank E. McKenzie for expressing human KMO. We also thank Diamond Light Source for access to MX beamlines.

Author Contributions N.S.S., F.G., D.L. and T.F.O. initiated the project, designed experiments, analysed data and wrote the manuscript; M.A. cloned purified and crystallized proteins and performed biochemical assays; C.L. crystallized proteins, collected and processed diffraction data; D.J.H. developed and analysed some of the biochemical assays; P.L. performed molecular modelling of L-KYN binding.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (<http://www.pdb.org>) under accession numbers 4J2W, 4J31, 4J33, 4J36 and 4J34. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S.S. (nigel.scrutton@manchester.ac.uk).

METHODS

Cloning, expression and purification of human full-length KMO. The gene encoding full-length human KMO was synthesized (GeneScript) and codon optimized for overexpression in mammalian cells. The gene was subcloned into the baculovirus transfer vector pAcGHLT-A–glutathione S-transferase as a NdeI–EcoRI fragment and transfected into Hi5 cells along with linearized baculovirus using the following sense and antisense primers: 5′-GGCATATGCATGGACAGCAGCGTGATCCAGCGGAAG-3′; 5′-CCC GAATTCCTACCGGTGATCAG GTTGCTG-3′. Hi5 cells (1.5×10^6) were infected with recombinant virus for 72 h at 28 °C. A WAVE Bioreactor System (GE Healthcare Life Sciences) was used to grow batches of 5-litres of cell culture. Cultured cells were lysed in 20 mM potassium phosphate buffer, pH 7.5, 10% glycerol, 0.5% *n*-dodecyl β -D-maltoside (DDM), 150 mM NaCl, 7 mM 2-mercaptoethanol and 50 μ M FAD supplemented with protease inhibitors (Sigma–Aldrich). Soluble lysate was incubated with 3 ml pre-equilibrated glutathione uniflow resin (Clontech). The resin was then packed and washed with buffer B (20 mM potassium phosphate buffer, pH 7.5, 10% glycerol, 0.012% DDM, 150 mM NaCl, 7 mM 2-mercaptoethanol, 50 μ M FAD). Fractions (0.5 ml) were eluted with buffer C (buffer B + 33 mM glutathione) and fractions containing KMO pooled, concentrated and loaded onto a Superdex 200 (10/30) size-exclusion chromatography column. Pure KMO was pooled and stored at –80 °C.

Cloning, expression and purification of *S. cerevisiae* KMO. A synthetic gene (BNA4, UniProtKB accession number P38169) encoding *S. cerevisiae* KMO was codon optimized for overexpression in *E. coli* (GeneScript). The gene was subcloned into pET15b and pET24b (Merck) as a NdeI–XbaI fragment for overexpression in *E. coli* strain BL21(DE3). A deletion variant of KMO (KMO(Δ 394)) and single-amino-acid variants of this deletion (KMO(Δ 394, R83A) and KMO(Δ 394, R83M)) were generated by site-directed mutagenesis (QuikChange, Stratagene, Agilent Technologies). Sense and antisense primers are shown in Supplementary Table 1. Amplification conditions were as follows: denaturation at 95 °C for 30 s, followed by 16 cycles of amplification at 95 °C (30 s), 55 °C (60 s) and 68 °C (7 min). PCR products were incubated for 1 h to digest methylated template DNA and then transformed into XL1-Blue supercompetent cells (Stratagene, Agilent Technologies). Transformed bacterial colonies were isolated and mutated genes were identified by DNA sequencing (Eurofins MWG Operon) to verify presence of desired, and lack of spurious changes, to the DNA sequence.

Transformed bacterial cells were grown overnight at 37 °C in lysogeny broth (LB; 0.75 l) containing 100 μ g ml^{–1} ampicillin. Cultures were grown at 37 °C to an attenuation of ~0.3, and the temperature was then reduced to 27 °C. Cells were induced (0.1 mM isopropyl- β -D-thiogalactopyranoside; IPTG) and grown (16–20 h) at 27 °C. Cells were lysed in buffer A (20 mM potassium phosphate buffer, pH 7.5, 10% glycerol, 300 mM NaCl, 50 μ M FAD) containing protease inhibitors. Soluble lysate was loaded onto a 5-ml HisTrap FF Ni Sepharose 6 Fast Flow column (GE Healthcare) pre-equilibrated with buffer A. Protein was eluted using a 50–250-mM imidazole linear gradient. Yellow fractions containing partially purified KMO were pooled and diluted ten times in buffer B to reduce salt and imidazole concentration. Protein was subsequently loaded onto a 10-ml blue Sepharose 6 fast flow column (GE Healthcare) equilibrated with buffer B. The protein was eluted using a linear gradient to 2.5 M NaCl prepared in buffer B. Fractions containing KMO were pooled and diluted tenfold in buffer B to reduce the NaCl concentration. Protein was concentrated and chromatographed using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with buffer C (25 mM ammonium acetate buffer, pH 7.0, 150 mM NaCl, 7 mM 2-mercaptoethanol). Pure fractions were collected and stored at –80 °C.

Incorporation of selenomethionine into *S. cerevisiae* KMO. Expression of selenomethionine-labelled protein was achieved by inhibiting methionine biosynthesis shortly before induction of KMO expression by adding high concentrations of isoleucine, leucine, phenylalanine, lysine and threonine to the cell culture. Transformed *E. coli* cells containing the *S. cerevisiae* KMO expression plasmid were grown in LB media at 37 °C to late exponential phase, collected and then resuspended in M9 minimal media and grown until mid-log phase. At that point lysine, phenylalanine, threonine (100 mg l^{–1} each) and selenomethionine, isoleucine, leucine and valine (50 mg l^{–1} each) were added to the culture and induced 15 min after addition of amino acids with 0.1 mM IPTG. The culture was grown for 12–16 h. Purification of selenomethionine-incorporated *S. cerevisiae* KMO was as described for conventional *S. cerevisiae* KMO.

Enzyme assays. Steady-state kinetic parameters for KMO were obtained by initial rate measurements of enzyme activity. KMO catalyses the NADPH-dependent hydroxylation of L-KYN to 3-HK and the enzymatic reaction can be monitored by following the decrease in absorbance of NADPH at 340 nm. Rate assays were carried out with 1-cm path-length quartz microcuvette at 37 °C for human KMO and 30 °C for *S. cerevisiae* KMO. Pure enzyme (0.2–1.0 μ M) was added to 200 μ l reaction buffer (20 mM potassium phosphate buffer, pH 8.0, 7 mM 2-mercaptoethanol) containing

different concentrations of NADPH and L-KYN (Sigma–Aldrich) and the time-dependent absorbance change at 340 nm was recorded using a UV-Vis Cary Eclipse spectrophotometer (Agilent Technologies). Assays at each reaction condition were performed in triplicate. Apparent Michaelis constants (K_m) for L-KYN and NADPH were determined by varying the concentration of the first substrate at a constant concentration of the second substrate and vice versa. Reaction data were fitted to the standard Michaelis–Menten equation using Origin Software (OriginLab).

KMO inhibition by UPF 648 was measured by monitoring the formation of the product 3-HK by C₁₈ reversed-phase HPLC. Enzyme (100 nM) was incubated at 37 °C in 1 ml reaction buffer (20 mM potassium phosphate buffer, pH 8.5, 7 mM 2-mercaptoethanol) with 2 mM NADPH, 500 μ M L-KYN and different concentrations (0–2 μ M) of UPF 468. The reaction was stopped with 8% trifluoroacetic acid added at different time points during the assay. Precipitated protein was removed by centrifugation, and the supernatant was analysed by C₁₈ reversed-phase HPLC equilibrated with 50 mM ammonium acetate, pH 3.0, containing 1% methanol and 0.1% heptanesulphonic acid; elution from the column was with 50 mM ammonium acetate, pH 4.5, containing 5% methanol and 0.5% heptanesulphonic acid, and monitored at 229 nm. 3-HK was quantified by comparison with a 3-HK calibration curve. Each experiment was performed in triplicate. Data were fitted to the Morrison equation:

$$\frac{v_0}{v_i} = 1 - \frac{([E] + [I] + K_i) - \sqrt{([E] + [I] + K_i)^2 - 4[E][I]}}{2[I]} \quad (1)$$

in which v_0 is the enzyme activity without inhibitor, v_i is the enzyme activity at ligand concentration I , $[E]$ is the enzyme concentration, $[I]$ is the inhibitor concentration and K_i is the inhibition constant.

Ligand-binding assays. Fluorescence-emission measurements were performed using a Cary Eclipse Fluorimeter (Agilent Technologies) to determine enzyme–ligand K_d for KMO ligands, exploiting ligand perturbation of flavin (enzyme-bound FAD) fluorescence. Excitation light was provided from a xenon light source and excitation and emission slit widths were 5 nm. The perturbation in FAD fluorescence emission was followed at 520 nm using an excitation wavelength of 450 nm and excitation spectra were recorded from 470 to 700 nm. L-KYN and UPF 648 were titrated into a 5- μ M solution of KMO (20 mM potassium phosphate buffer, pH 8, 50 mM NaCl, 7 mM 2-mercaptoethanol) at 25 °C. K_d was calculated by fitting data for the difference in fluorescence emission measured for KMO–ligand complexes and free KMO at 520 nm divided by the maximum emission versus the ligand concentration (equation (2)). This equation is a variation of Morrison equation used in equation (1) and is used for tight-binding ligands or when K_d values are similar in magnitude to the concentration of enzyme used, in which case substantial amounts of both enzyme and ligand are consumed in forming the EX complex as the titration progresses. All data fitting was done using Origin software (OriginLab).

$$\frac{E_{\text{obs}}}{E_{\text{max}}} = 1 - \frac{([E] + X + K_d) - \sqrt{([E] + X + K_d)^2 - 4[E]X}}{2[E]} \quad (2)$$

In equation (2), E_{obs} represents the fluorescence-emission difference at each ligand concentration, X ; E_{max} is the maximum emission change at saturating ligand concentrations, $[E]$ is the enzyme concentration and K_d the dissociation constant for the ligand bound to the enzyme.

Hydrogen peroxide quantification. Measurements to observe the formation of hydrogen peroxide from the Δ KMO-394–UPF 648 inhibitor complex were performed using a horseradish peroxidase (HRP) assay. Steady-state assays were performed as described above to calculate the NADPH consumption rate in the presence and absence of an excess of UPF 648 (30 μ M). KMO (5 μ M) was incubated with 150 μ M NADPH, 150 μ M L-KYN, 400 μ M o-dianisidine and ~5 units HRP in the presence and absence of excess UPF 648. The rate of oxidation of o-dianisidine by H₂O₂, catalysed by HRP, was monitored at 440 nm ($\Delta\epsilon_{440\text{ nm}} = 11.600\text{ M}^{-1}\text{ cm}^{-1}$). The rate of hydrogen peroxide production was compared to the rate of NADPH consumption.

Crystallization of *S. cerevisiae* KMO. Initial crystals of *S. cerevisiae* KMO (KMO(Δ 396) and KMO(Δ 396)–Se) were obtained by mixing 200 nl of 14 mg ml^{–1} protein in 20 mM ammonium acetate, pH 7.0, 150 mM NaCl and 7 mM 2-mercaptoethanol (buffer A) with 200 nl of a reservoir solution containing 0.1 M sodium acetate, pH 5.5, and 35% isopropanol. A second, more readily reproducible, crystal form was obtained with Δ KMO-394 and Δ KMO-394–HIS[–]. In both cases crystals were grown by mixing 200 nl of protein (buffer A) with 200 nl of a reservoir solution containing 0.1 M imidazole, pH 7.8, and 11% w/v polyethylene glycol at 277 K. The UPF complex (Δ KMO-394–UPF 648) was also obtained as described above; however, before setting the trays the protein was

pre-incubated with 1 mM UPF 648 for ~30 min. All trays were incubated at 277 K, with crystals forming over a period of ~72 h.

Diffraction data processing, structure determination and refinement. X-ray diffraction data were collected at Diamond Light Source and subsequently integrated and scaled using the program XDS³¹. Initial phases were obtained from a single-wavelength anomalous dispersion data set (S1) collected at the selenium edge. Selenium sites were located using PHENIX AutoSol³² yielding an electron density map that could be auto traced using PHENIX AutoBuild³³. The resulting model was completed through iterative rounds of rebuilding in Coot³⁴ and refinement in PHENIX³⁵. All subsequent structures were solved by molecular replacement in Phaser³⁶ using this initial selenomethionine-derived structure as the template. Structure validation with MolProbity³⁷ was integrated as part of the iterative rebuild and refinement procedure.

KMO–L-KYN modelling. NAMD software³⁸ was used to perform all molecular dynamics simulations of the KMO–L-KYN complex. Topology and parameters files for substrate and FAD were obtained using Antechamber program³⁹ using AM1-BCC charges⁴⁰. The complex model was then immersed in a periodic water box (TIP3) and neutralized by adding Na⁺ ions. Several cycles of minimizations (steepest descent, 10,000 steps) and molecular dynamics simulations (50 K, 20 ps) were performed to equilibrate the model (backbone protein atoms were kept fixed). Then molecular dynamics simulations were performed (310 K, 1 ns) at a time step of 2 ps, with the protein backbone restrained on the X-ray structure

conformation. Individual snapshots were extracted and minimized to assess the KMO–L-KYN interactions.

31. Kabsch, W. *XDS. Acta Crystallogr. D* **66**, 125–132 (2010).
32. McCoy, A. J., Storoni, L. C. & Read, R. J. Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr. D* **60**, 1220–1228 (2004).
33. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
34. Emsley, P. & Cowtan, K. *Coot: model-building tools for molecular graphics. Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. Adams, P. D. *et al.* *PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D* **66**, 213–221 (2010).
36. McCoy, A. J. *et al.* *Phaser crystallographic software. J. Appl. Crystallogr.* **40**, 658–674 (2007).
37. Chen, V. B. *et al.* *MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D* **66**, 12–21 (2010).
38. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
39. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
40. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).

CORRIGENDUM

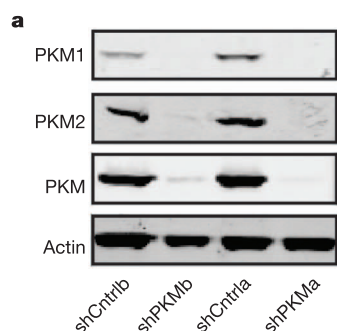
doi:10.1038/nature12075

Corrigendum: Serine is a natural ligand and allosteric activator of pyruvate kinase M2

Barbara Chaneton, Petra Hillmann, Liang Zheng, Agnès C. L. Martin, Oliver D. K. Maddocks, Achuthanunni Chokkathukalam, Joseph E. Coyle, Andris Jankevics, Finn P. Holding, Karen H. Vousden, Christian Frezza, Marc O'Reilly & Eyal Gottlieb

Nature **491**, 458–462 (2012); doi:10.1038/nature11540

In Fig. 1a of this Letter, the western blot originally published in the PKM2 panel was a re-probe of the blot used in the PKM1 panel and the signal picked up was from the first antibody (PKM1) and not from PKM2. We have repeated the experiment on three different membranes with the antibodies used in the original paper for PKM1, PKM2 and total PKM, using actin as the loading control (Fig. 1 of this Corrigendum shows the corrected Fig. 1a). We apologize for this inadvertent error. None of the conclusions of the original Letter are altered.



CORRIGENDUM

doi:10.1038/nature12078

Corrigendum: Mitochondrial Atp1f1 regulates haem synthesis in developing erythroblasts

Dhvanit I. Shah, Naoko Takahashi-Makise, Jeffrey D. Cooney, Liangtao Li, Iman J. Schultz, Eric L. Pierce, Anupama Narla, Alexandra Seguin, Shilpa M. Hattangadi, Amy E. Medlock, Nathaniel B. Langer, Tamara A. Dailey, Slater N. Hurst, Danilo Faccenda, Jessica M. Wiwczar, Spencer K. Heggors, Guillaume Vogin, Wen Chen, Caiyong Chen, Dean R. Campagna, Carlo Brugnara, Yi Zhou, Benjamin L. Ebert, Nika N. Danial, Mark D. Fleming, Diane M. Ward, Michelangelo Campanella, Harry A. Dailey, Jerry Kaplan & Barry H. Paw

Nature **491**, 608–612 (2012); doi:10.1038/nature11536

On page 610 of this Letter, the overview comments concerning the action of ATP synthase should read: “Under physiological conditions, the F_1F_0 -ATP synthase generates ATP by consuming the proton motive force existing across the inner mitochondrial membrane, transporting H^+ ions from the intermembrane space into the mitochondrial matrix^{8,15}. When respiration is inhibited or the mitochondrial membrane potential ($\Delta\Psi_m$) is decreased, the F_1F_0 -ATP synthase reverses the direction of H^+ transport, dissipating ATP in the attempt to maintain the proton motive force^{8,15}. Under these conditions, Atp1f1 inhibits the reversal of the F_1F_0 -ATP synthase to decrease ATP hydrolysis, thereby counteracting the maintenance of the mitochondrial membrane potential ($\Delta\Psi_m$) and proton motive force⁸”. This rewording has no effect on the data, the scientific interpretation and the resulting model presented, and has been corrected in the HTML and PDF versions of the original paper.

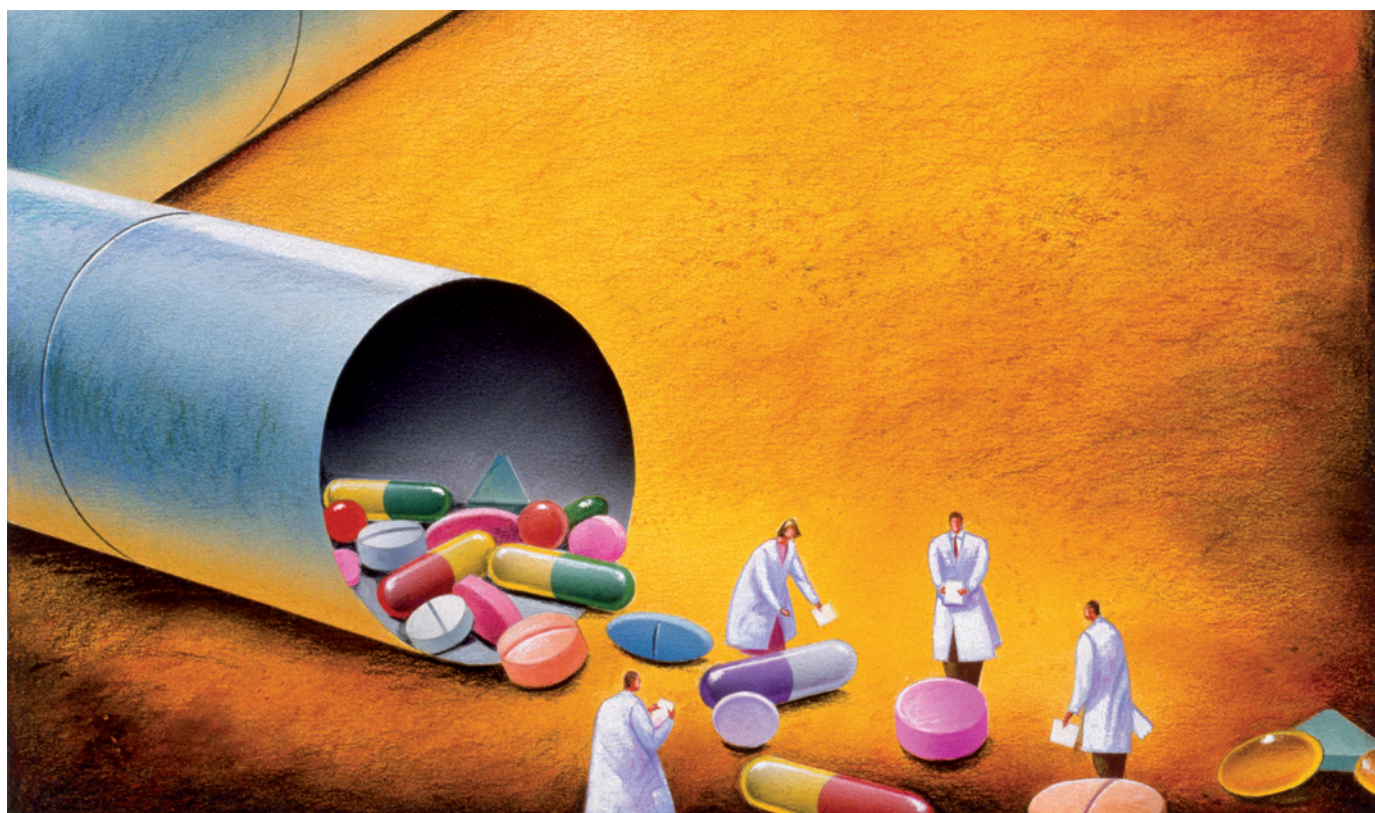
CAREERS

TURNING POINT Engineer takes up structural biology and cancer research **p.389**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com

IMAGES.COM/CORBIS



REGULATORY SCIENCE

Researchers in the pipeline

Prospects are starting to look bright in the growing field of science intended to aid regulation of food safety and drug development.

BY AMBER DANCE

When the Deepwater Horizon rig exploded in April 2010, releasing nearly 5 million barrels of oil into the Gulf of Mexico, fisheries were closed in more than one-third of US Gulf waters. Scientists had to test hundreds of samples of seafood for hydrocarbon contamination, but existing chemical tests took a week to yield results.

Researchers at the US Food and Drug Administration (FDA) and other agencies joined forces to develop a speedier assay that would let them rate seafood as safe or unsafe within two days. By refining methods that already existed, the agencies were able to get a

new test up and running by July 2010. Fisheries reopened and seafood was back on menus weeks earlier than would have been possible with the original test. Such quick, collaborative problem-solving exemplifies the benefits of regulatory science done not only in government agencies, but also in industry and academia.

Regulatory science is “not a new thing, it’s just a new term”, says Erin Wilhelm, project director at the Center for Excellence in Regulatory Science and Innovation (CERSI) in Washington DC, launched in 2011 by the FDA and Georgetown University. “It’s really an umbrella term for all different kinds of science that impact drug development or device development or food science or tobacco regulation.”

People wanting to work in regulatory science mostly need expertise in a relevant branch of science, but focused regulatory training is also available (see ‘Learning the ropes’). Jobs may involve reviewing applications for product approval, as well as a variety of research topics, such as identifying biomarkers for drug activity, gauging risks to public health during the drug-approval process or tracking a drug’s effects after it has reached the market. Companies and industry groups that need to gain approval for their products are seeking team members who understand the regulatory world. Jobs are available in many countries, especially in the United States, where the field has garnered interest in government and other sectors. ►

► Most people distinguish between regulatory science and regulatory affairs, which focuses on making sure that regulations are followed. “We’re talking real science,” says Hans-Georg Eichler, senior medical officer at the European Medicines Agency (EMA) in London. People who do regulatory science have diverse backgrounds, ranging from biochemistry to statistics. And they tend to think of themselves as statisticians or biochemists, not ‘regulatory scientists’, he says.

“It’s not been an area of high profile,” notes John Burris, president of the Burroughs Wellcome Fund in Research Triangle Park, North Carolina, which funds biomedical research. “People aren’t winning Nobel prizes in it.” Yet the results of regulatory science have the potential to make enormous impact, he says. For example, a job in basic research identifying biomarkers for disease could help scientists running clinical trials to identify who is responding well to treatment. And by testing drugs in model systems, researchers may be able to learn about the efficacy and side effects of a compound before taking on the cost and risk of human trials.

Model systems currently under development by regulatory scientists include the ‘human on a chip,’ which simulates ten different organs for drug-safety testing. Another is the virtual family — anatomically correct digital models of male and female adults and children, based on body scans of volunteers and used in computer simulations. The virtual family is already being applied to predict body-temperature changes in the tissues surrounding implanted devices when people are exposed to radio waves or microwaves.

OPPORTUNITIES ARISING

Governments and institutions have started to take note of the field. The Burroughs Wellcome Fund is evaluating applications for its first Innovation in Regulatory Science Awards — grants for academics addressing regulatory topics — which it initiated to kick-start investment in the area. It has received applications from about 60 faculty members in North America, and it plans to endow at least five projects with up to US\$500,000 each over five years. If the programme is successful, Burris hopes to make it an annual call for proposals.

The FDA does not offer many research grants, although it partnered with the US National Institutes of Health in 2010 to offer a total of US\$6.75 million for a few three-year projects in regulatory areas such as disease biomarkers. Last year, it also issued a call for external contractors to conduct regulatory research.

In Japan, the Pharmaceuticals and Medical Devices Agency established an Office of Regulatory Science in 2009. The Ministry of Health, Labour and Welfare offered ¥1.2 billion (US\$12 million) in grants in the regulatory field in 2012.

In Europe, the Innovative Medicines Initiative (IMI) in Brussels supports public-private

collaborations on topics such as biomarkers and disease models in the hope of improving the drug-development process.

The blooming interest in regulatory science has not necessarily translated into job opportunities at the FDA or EMA, however. Eichler says that the EMA’s job market was growing rapidly until 2009, when the economy began its downward spiral and hiring stalled. IMI executive director Michel Goldman, by contrast, sees opportunities worldwide, including in Europe, Japan and China. And a 2012 review by the non-profit Partnership for Public Service in Washington DC found that the FDA hired 2,221 people in 2010, up from 817 in 2007. However, many positions are temporary. And although FDA chief scientist Jesse Goodman anticipates hiring more people as the older generation of employees retires, this year’s federal budget cuts have the potential to hit the agency hard. The FDA spends the majority of its money on staff and related costs.

Although no forced unpaid leave or lay-offs are expected, the budget cuts could lead to delayed hiring and reduced travel and overtime.

PETRI DISHES AND PAPERWORK

A regulatory agency is the obvious — although not the only — place to start a career in regulatory science. Some FDA employees have left experimentation behind, whereas others split their time between research projects and reviewing applications for product approvals.

In some ways, working in regulation is like being a principal investigator, says Candice Jongsma, a regulatory-science fellow at the FDA Center for Tobacco Products in Rockville, Maryland. She may not do the experiments herself, but she reviews data, asks questions and makes recommendations. One of her projects is to compare the lists of ingredients in new and existing tobacco products and determine whether the new recipe raises health questions that will need to be addressed before approval.

TRAINING PROGRAMMES

Learning the ropes

Training in regulatory science has tended to be an on-the-job apprenticeship, with the major requirement a solid background in science, but targeted education is also available.

Fellowships can help researchers to explore the regulatory side of science before applying for a permanent position. Candice Jongsma is among the first Tobacco Regulatory Science Fellows at the US Food and Drug Administration (FDA) Center for Tobacco Products in Rockville, Maryland; her one-year programme is co-sponsored by the US Institute of Medicine. Jongsma applies her chemistry knowledge to look for safety concerns in the ingredients of new products, and her fellowship includes training courses and personal meetings with the secretary of Health and Human Services and the commissioner of the FDA. The FDA also sponsors a two-year Commissioner’s Fellowship, which may open the door to further work at the agency after they have completed the programme.

Meanwhile, educational programmes focused on regulation are on the rise. They are often geared towards working professionals, and involve weekend or evening classes and online learning. Some curricula concentrate on science and drug development; others focus more on laws and regulations, or pricing and business decision-making, says Frances Richmond, director of the International Center for Regulatory Science at the University of Southern California in Los Angeles, which offers master’s and doctoral degrees.

Master’s programmes are in development, and are expected to open this year, at Georgetown University in Washington DC; the University of Maryland in Baltimore, as an online course; and the University of Pennsylvania in Philadelphia. In the Maryland programme, for example, students will get a grounding in regulatory affairs, drug discovery and development, clinical research and surveillance to track the safety of a drug after it has been approved.

In Europe, students can enrol in short courses lasting a few days, or year-long master’s programmes. Many focus on regulatory affairs and some on regulatory science. The Innovative Medicines Initiative (IMI) in Brussels runs the postgraduate PharmaTrain programme, which includes short classes and master’s-level training in drug development. Its classes are available in several locations across the continent. The IMI’s European programme in Pharmacovigilance and Pharmacoeconomics, called Eu2P, offers online training at the certificate, master’s and PhD level.

The Japanese Pharmaceuticals and Medical Devices Agency has united with a dozen universities to allow its employees to earn graduate degrees and to let university graduate students conduct their thesis research at the agency.

“Talent is hard to find,” says Richmond. “A person with good science skills and a degree in regulatory science generally has no problems getting a good job.” **A.D.**

Ernest Kwegyir-Afful was a neuroscience postdoc when he applied for a two-year fellowship at the FDA in the hope of getting to conduct research that would have an immediate impact. At the FDA's Center for Food Safety and Applied Nutrition in College Park, Maryland, Kwegyir-Afful worked on estimating the risk that small quantities of food allergens might endanger sensi-

"It's a field that you would enjoy if you don't want to think one-dimensionally."

He came up with a computational method for risk analysis that was closely tailored to allergens. After finishing his fellowship last September, Kwegyir-Afful accepted a permanent position at the centre, doing similar work. Regulatory science is "a field that you would enjoy if you don't want to think one-dimensionally," he says, noting that he works with experts in topics such as toxicology or law. Communication skills and a team-oriented approach are essential, says Frances Richmond, director of the International Center for Regulatory Science at the University of Southern California in Los Angeles.

Regulatory-strategy units at pharmaceutical companies are good places to explore the regulatory side of industry, says Robert Meyer, who directed the FDA Office of Drug Evaluation before becoming vice-president of global regulatory strategy for Merck at its campus in Upper Gwynedd, Pennsylvania. Company regulatory scientists would do well to understand the relevant laws for FDA reviewers, which can help company scientists to do the right experiments and to offer the FDA the data it needs to make decisions, says Meyer. That is especially true for people working at small companies, because they participate in many aspects of product development, says James Polli, co-principal investigator of an FDA CERSI at the University of Maryland in Baltimore.

Academic scientists, too, can address regulatory questions or advise the FDA. Subha Madhavan, a bioinformatician at Georgetown University, is building databases that will be used by FDA scientists looking to understand why people respond to drugs in different ways, and how vaccines might be linked to autoimmune disease. She regularly meets FDA staff to explore their needs, and expects to publish the work.

"Regulatory science is the field of the future, if you will, in terms of drug development and device development," says Wilhelm. "At some point, it will mean something to say, 'I am a regulatory scientist.'" ■

Amber Dance is a freelance writer in Los Angeles, California.

TURNING POINT

Adam de la Zerda

Adam de la Zerda was on track to become an electrical engineer when a personal loss prompted him to switch to cancer biology. Now a structural biologist at Stanford University in California, de la Zerda has developed a technique for imaging tumour cells.

Why did you take up cancer biology?

In my undergraduate programme at the Technion Israel Institute of Technology in Haifa, I focused on computer science, engineering and physics. When I started my PhD at Stanford, I was working on experimental quantum physics. But during winter break in my first year, a good friend passed away from cancer. That was a true turning point for me. I started examining how to apply my background to cancer research.

Was it a difficult leap?

No. Stanford has a flexible structure that can accommodate such transitions. I could have an adviser from anywhere in the university; I found an amazing one in Sanjiv Gambhir, who was working on molecular imaging. I knew that early detection of cancerous tissue could have prevented many cancer deaths, including my friend's. Gambhir agreed to let me join his lab without a biology background.

How did you persuade him?

I think my motivation convinced him, but I had also read a lot about his imaging work. I was able to explain why I was so excited about the prospects for direct applications. I launched a project to convert light into sound so we could look into tissue and find out whether cells are cancerous. My programming skills helped me to create our first real imaging unit using lasers and ultrasound systems.

Why did you pursue a postdoc in chemistry?

As my PhD ended, I realized that much of what we do as biologists relies on chemistry. I had never taken a chemistry class, and wanted to study it. My dad has a PhD in chemistry, which also helped me to appreciate the subject. Luckily, Carolyn Bertozzi, a chemist at the University of California, Berkeley, accepted me as the first engineer in her lab. My knowledge of chemistry was at a first-year-undergraduate level, but I learned a lot, often from undergraduates.

Did you plan to do a one-year postdoc?

No — I had a three-year fellowship. Some people want to stay in their field and crank out papers; I wanted to expand my toolkit with organic chemistry. After a year, I saw an opportunity to apply for a faculty position in



structural biology at Stanford. I was in no rush to leave the postdoc, but I felt ready. I got the position and started last autumn.

How did you approach the interview?

It is important to have an interesting story to tell when you apply. Given the tough job market, hiring committees are risk-averse. They don't want to hear, "I want to solve X problem." They want to hear, "This is exactly how I'm going to solve X problem." I feel fortunate that the search committee was willing to accept me, whereas others may have said I was at too early a stage.

Last year, you won the Dale F. Frey Award for Breakthrough Scientists. How will it help you?

It is awarded by the Damon Runyon Cancer Research Foundation in New York, which contributes 100% of the donations it receives to research that will affect patients. Through the foundation, I have met venture capitalists whom I am trying to bring to research brainstorming sessions. I want to work out how best to translate my group's research and patents into commercial medical devices. I want the harsh feedback at the start of a project to help us to make sure that our technologies will help patients. We are willing to take big risks, but we want to make sure that there are big benefits.

Should engineers explore life sciences?

Yes. In the past year I have been teaching a class that exposes engineering students to biology; 300 students have taken it. I show them how engineering tools can solve problems in biology and medicine, especially in cancer and cardiovascular disease. Several engineering students have joined immunology and cardiology labs. That makes me extremely proud. ■

INTERVIEW BY VIRGINIA GEWIN

SURVIVORS AND SAVIOURS

The ultimate challenge.

BY PHILIP T. STARKS

The airlocks close behind me, but I can't hear them. My ears register nothing but the radio and my breathing. White noise, the rhythmic acceptance and expulsion of recycled air, and my own inner dialogue keep me company as I take my first step. The ground compresses beneath my foot, but my protective gear mutes the full experience.

I was selected for this mission not because I'm the oldest or youngest, weakest or strongest, smartest or dimmest, but because I'm expendable. When you're a dead-end vehicle in an evolutionary war, your job is to die first. I've been awarded this step because I'm sterile.

Sterile not only describes me, but also the environment that I've survived in. I won't say, 'lived in', although the others seem hopeful enough to equate survival with living. I don't. Up until now, I've been surviving. However, with each step taking me farther from the enclave into the real world, I feel more alive.

And I'm not alone. Out here, the world is teeming with life. Some teams collaborate, some hold ground in extended battle and others triumph. The skeletal remains splintering beneath my feet, serving as coral-like scaffolding for microbial masters, remind me that when one team triumphs, another fails. I might be advancing, but my team is losing.

Many teams have lost, but I suppose the majority have survived. It's just my own bias that elevates birds over biofilm. But I find no comfort in these thoughts. So, while ploughing through this microbial minefield, I focus on stories of long extinct beasts: of mammals, birds, reptiles and insects. I recite parables explaining how the loss of honeybees was an ignored bioindicator of a New Age, and how financial and political interests handcuffed a timely response. I fill my mind with these details to shield me from my fears.

My mission is to find survivors, but the unspoken hope is to find females. That hope isn't buoyed by data or experience. Of almost 10,000 embryos, only 1,500 survived. Even this pathetic number stressed our pool of viable uteruses, forcing us to transfer as many as three to the younger soldiers. Most failed to implant, or perished shortly there-



after. Our hopes were pinned on 350 fetuses, 280 of which survived birth; 165 enjoyed their ninth birthday, and prior to their tenth, we released them. Only 68 were female.

Now I'm searching for at least one. I understood that this trip would be long, and that the first steps would reveal our past failures. Despite these warnings, I didn't anticipate the boneyard I'd trample. The beauty of the growth encompassing them notwithstanding, these were my kin. Still, the horror of the eviscerated, cleanly consumed bodies was infinitely preferable to what I feared came next.

I pass through the final gateway, frightened but hopeful, but ultimately unprepared. My fears, developed and nourished in isolation, paled against this sight. Extinction nestles into one's arms more easily than a dead child. And I can't possibly hold all of the dead children.

Most of the food and water packets littering the ground are untouched. I am pleased, actually, by the advanced decomposition:

most died immediately. Some of the bones, however, hold flesh. These children had survived, but not

long enough to rejoin us. I mourn them, mutter a silent prayer, and wish they were never born.

I begin tallying the dead, then stop and radio in: "I count at least 150, I'm coming back." I turn down the volume and don't wait for an answer. We fertilized the eggs, developed the embryos and reared the young. We took the survivors and released them, unprotected, into this world. We killed them: I just couldn't count every last dead child.

If I had, my count would have fallen short. My walk back to my sterile tomb was not solitary, and before I reached the airlock I felt a delicate hand tug my gloved finger.

"Don't leave me."

She is beautiful: eleven years old, fit, and without protective gear. Her words are stronger than she knows: with them, she has sealed both our fates. I am overcome with joy and grief: I am thrilled she breathes, but I wish I was dead.

"I won't leave you," I promised her. "Where are your friends?"

"Gone," she said. I understood. She was the lone survivor of our challenge test. She has endured our abuse, and now I'll deliver her back to her abusers.

"So many people are looking forward to meeting you!" I confess to her the truth, but not its depth. I know that delivering her to the airlock will kill her freedom. She has survived, but only to be an egg donor. The goal of our research has always been to produce a generation that could thrive without protection. If she carries alleles that protect her against extinction-level pathogens, we need them.

I bring her to the airlock, call in the news and lie to her. "I'm going to step aside and protect the site, so I know you make it in." As she enters our crypt, I remove my mask, remember those we saved and mourn those we killed. As a dead-end vehicle, I am happy to reach my end. I breathe deeply, taste the Age of Microbes, and convince myself that the Age of Man is not over. ■

Philip T. Starks is an associate professor of biology at Tufts University. His research focuses mostly on the behaviour of social insects, but he has recently been exploring microbial communities.

➔ **NATURE.COM**
Follow Futures:
@NatureFutures
f go.nature.com/mtoodm